ΤΕΙ ΔΥΤΙΚΗΣ ΕΛΛΑΔΑΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ

ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

«ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ»

ΚΑΛΥΒΑ ΠΑΝΑΓΙΩΤΑ

ΛΑΖΑΡΟΥ ΜΑΡΙΕΛΕΝΑ

ΜΥΛΩΝΑ ΔΙΟΝΥΣΙΑ

ΕΠΟΠΤΕΥΩΝ ΚΑΘΗΓΗΤΗΣ: ΔΡ. ΒΑΣΙΛΙΚΗ ΚΑΡΙΩΤΗ

ΕΙΣΗΓΗΤΗΣ: ΦΩΤΕΙΝΟΠΟΥΛΟΣ ΑΝΑΣΤΑΣΙΟΣ- ΜΙΧΑΗΛ

ΠΑΤΡΑ, 2015

ΠΡΟΛΟΓΟΣ

Η Στατιστική είναι μία μεθοδική μαθηματική, παλαιότερα τεχνική και σήμερα επιστήμη που επιχειρεί να εξαγάγει έγκυρη γνώση χρησιμοποιώντας εμπειρικά δεδομένα παρατήρησης, ή πειράματος. Κύριο αντικείμενο έρευνας και μελέτης της Στατιστικής είναι η συλλογή, ταξινόμηση, επεξεργασία, παρουσίαση, ανάλυση και ερμηνεία διαφόρων δεδομένων με απώτερο στόχο την εξαγωγή ασφαλών συμπερασμάτων για λήψη ορθών αποφάσεων. Πρόκειται για σημαντική επιστήμη της οποίας οι εφαρμογές έχουν ευρύτατο πεδίο στη διοικητική, τις επιχειρήσεις, καθώς και στιςθετικές και συμπεριφορικές ή Κοινωνικές επιστήμες

Με τον έλεγχο υποθέσεων προσπαθούμε να διαπιστώσουμε εάν τα δεδομένα του δείγματος υποστηρίζουν την υπόθεση ότι η παράμετρος του πληθυσμού έχει μια συγκεκριμένη τιμή.

Στους στατιστικούς ελέγχους υποθέσεων:

• Επιθυμούμε να ελέγξουμε αν μία ή περισσότερες παράμετροι (π.χ. μέση τιμή μ) ενός πληθυσμού ικανοποιούν μια βασική υπόθεση (π.χ. μ = 100) έναντι μιας εναλλακτικής υπόθεσης (π.χ. μ > 100).

• Σχεδόν πάντοτε δεν είμαστε σε θέση να καταγράψουμε όλον τον πληθυσμό (ώστε να αποφανθούμε με βεβαιότητα για το αν ισχύει η βασική υπόθεση) οπότε αρκούμαστε σε ένα τυχαίο δείγμα από αυτόν.

• Με βάση αυτό το τυχαίο δείγμα θέλουμε να πάρουμε μια απόφαση: να απορρίψουμε ή όχι ότι ισχύει η βασική υπόθεση.

Συνήθως εργαζόμαστε ως εξής:

• Κατασκευάζουμε κάποια συνάρτηση του δείγματος (στατιστική συνάρτηση) η οποία, όταν ισχύει η βασική υπόθεση, ακολουθεί μια συγκεκριμένη (γνωστή) κατανομή. Ενώ όταν ισχύει η εναλλακτική υπόθεση, λαμβάνει «ακραίες» τιμές (πολύ «μεγάλες» ή πολύ «μικρές»).

• Αν, με βάση το τ.δ. που πήραμε, αυτή η στατιστική συνάρτηση λάβει κάποια «ακραία» τιμή τότε απορρίπτουμε την βασική υπόθεση. Το πότε μια τιμή θεωρείται «ακραία» ώστε να απορρίψουμε την βασική υπόθεση εξαρτάται από το «επίπεδο σημαντικότητας» που έχουμε προαποφασίσει.

ΠΕΡΙΛΗΨΗ

Η παρούσα πτυχιακή εργασία πραγματεύεται τον έλεγχο υποθέσεων. Στο πρώτο κεφάλαιο αναφέρονται ορισμένα εισαγωγικά στοιχεία σχετικά με τον έλεγχο στατιστικών υποθέσεων, τη διατύπωση και τα σφάλματα του.

Στο δεύτερο κεφάλαιο αναλύονται οι έλεγχοι υποθέσεων σχετικά με ένα πληθυσμό με τη μέση τιμή, τη διασπορά, την αναλογία.

Στο τρίτο κεφάλαιο τίθενται οι εφαρμογές των ελέγχων υποθέσεων, τα στοιχεία εφαρμογής τους, ο ρόλος τους στην οικονομική επιστήμη και το ευρύτερο πεδίο εφαρμογής.

Στο τέταρτο κεφάλαιο παρουσιάζεται η μελέτη περίπτωσης που εφαρμόστηκε σε ιδιωτική εταιρεία 100 υπαλλήλων με σκοπό τη συγκέντρωση στοιχείων για τα χαρακτηριστικά τους μέσω του στατιστικού προγράμματος minitab.

Τέλος, παρατίθενται οι βιβλιογραφικές αναφορές που χρησιμοποιήθηκαν και το παράρτημα με τα πλήρη δεδομένα που χρησιμοποιήθηκαν στη μελέτη περίπτωσης.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ

Ο όρος της στατιστικής υπόθεσης, ή αλλιώς statistical hypothesis, όπως αναφέρεται στην διεθνή βιβλιογραφία, στηρίζεται στην κατανομή μιας ή και περισσοτέρων τυχαίων μεταβλητών.

Η διενέργεια ελέγχου υποθέσεων για την μέση τιμή διακρίνεται σε δύο είδη ελέγχων, ανάλογα με το αν είναι η τυπική απόκλιση είναι γνωστή ή όχι.

Έλεγχος υπόθεσης μπορεί να διενεργηθεί όταν υπάρχουν ομοειδής εκτιμήσεις για μία παράμετρο, είναι ο έλεγχος για την διαφορά των αναλογιών.

Η εφαρμογή της επιστήμης της Στατιστικής και συγκεκριμένα του ελέγχου των υποθέσεων βρίσκει μεγάλο αντίκτυπο σε πλήθος τομέων και άλλων επιστημών. Η στατιστική συμπερασματολογία εφαρμόζεται σήμερα στην επιστήμη της Ιατρικής, της Βιολογίας, την Φυσική, την Γενετική, την Αστρονομία, την Μετεωρολογία, την Γεωργία, την Βιομηχανία, κ.λπ.

Ο έλεγχος της θεωρίας συνίσταται στην αντιπαραβολή των ποιοτικών (λογικών) συμπερασμάτων της οικονομικής ανάλυσης με τα πραγματικά δεδομένα, προκειμένου να διαπιστωθεί ο βαθμός που αυτά ερμηνεύουν την πραγματικότητα. Ο έλεγχος αυτός επιτυγχάνεται με τη βοήθεια μεθόδων, όπως είναι οι έλεγχοι των υποθέσεων που αναπτύσσονται μέσα από την επιστήμη της στατιστικής συμπερασματολογίας.

Ο Στατιστικός Έλεγχος Ποιότητας αποτελεί ίσως την πιο γνωστή μέθοδο με την οποία γίνεται ο έλεγχος της λειτουργίας της παραγωγής μίας επιχείρησης, καθώς και με ποιο τρόπο μπορεί να επιτευχθεί βελτίωση της ποιότητας των παραγόμενων προϊόντων. Ένας από τους κυριότερους στόχους του στατιστικού ελέγχου ποιότητας είναι η έγκαιρη διαπίστωση των προϊόντων που δεν πληρούν τια απαραίτητες προδιαγραφές των προϊόντων που παράγει η επιχείρηση. Κατά την διαδικασία του ελέγχου, όταν διαπιστώνεται στην παραγωγική διαδικασία ένα προϊόν το οποίο δεν συμμορφώνεται με τις προδιαγραφές που έχουν τεθεί από την διοίκηση, τότε ξεκινάει μία σειρά διαδικασιών για την λήψη διορθωτικών ενεργειών και την εξακρίβωση των αιτιών που προκαλούν τις μη συμμορφώσεις, με στόχο πάντοτε την διατήρηση της ποιότητας των προϊόντων.

ΚΕΦΑΛΑΙΟ ΠΡΩΤΟ
ΕΙΣΑΓΩΓΙΚΑ ΣΤΟΙΧΕΙΑ ΓΙΑ ΤΟΝ ΕΛΕΓΧΟ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ

1.1 ΕΙΣΑΓΩΓΗ

Το παρών κεφάλαιο αποτελεί το εισαγωγικό κομμάτι για τον έλεγχο των υποθέσεων. Αρχικά δίνονται κάποιοι ορισμοί που είναι απαραίτητοι για την συνέχεια της παρούσας πτυχιακής εργασίας, ενώ στην συνέχεια προσδιορίζεται ο τρόπος με τον οποίο διατυπώνονται και ελέγχονται οι στατιστικές υποθέσεις.

1.2 ΣΤΑΤΙΣΤΙΚΗ ΥΠΟΘΕΣΗ

Ουσιαστικά, μία στατιστική υπόθεση, μπορεί να είναι μία υπόθεση που να περιγράφει την συμπεριφορά ορισμένων τυχαίων μεταβλητών, μέσα από ένα πλήθος παρατηρήσεων.[1]

Για τον έλεγχο των στατιστικών υποθέσεων, απατούνται δύο υποθέσεις. Η πρώτη υπόθεση που δημιουργείται αποκαλείται μηδενική υπόθεση (null hypothesis) και προσδιορίζει ότι, μία παράμετρος θ μπορεί να λάβει μία συγκεκριμένη τιμή θ_0. Η υπόθεση αυτή συμβολίζεται με H_0 και η μορφή που λαμβάνει είναι η εξής: $H_0{:}\theta=\theta_0$ (δηλαδή η παράμετρος θ ισούται με την συγκεκριμένη τιμή θ_0).

Η δεύτερη υπόθεση, η αλλιώς εναλλακτική υπόθεση (alternative hypothesis) όπως αποκαλείται στην επιστήμη της στατιστικής συμβολίζεται ως H_1. Προτού διατυπωθούν οι μορφές που μπορεί να λάβει η εναλλακτική υπόθεση σε μία έρευνα, θα πρέπει να σημειωθεί ότι, το είδος του ελέγχου που θα ακολουθηθεί, καθορίζεται κάθε φορά από την μορφή που έχει δοθεί στην εναλλακτική υπόθεση. Οι έλεγχοι που μπορούν να πραγματοποιηθούν είναι ο μονόπλευρος έλεγχος και ο δίπλευρος έλεγχος.

[1] http://www.aua.gr/gpapadopoulos/files/hypoth-tests-4.pdf

Στην πρώτη περίπτωση, στον μονόπλευρο έλεγχο, οι διαφορές που μας ενδιαφέρουν κάθε φορά εστιάζουν προς το καθορισμένο από αυτόν άκρο της κατανομής. Στη περίπτωση του δίπλευρου ελέγχου, ελέγχονται οι διαφορές της μέσης τιμής του δείγματος και δεξιά και αριστερά από την μέση τιμή της μηδενικής υπόθεσης.

Συνεχίζοντας στην διατύπωση της εναλλακτικής υπόθεσης, οι μορφές που μπορεί να λάβει η H_1 είναι οι ακόλουθες:

- H_1: $\theta \neq \theta_0$, δηλαδή η παράμετρος θ θα είναι μικρότερη ή μεγαλύτερη από την τιμή θ_0. Στην συγκεκριμένη περίπτωση ο έλεγχος που πραγματοποιείται είναι ο δίπλευρος. Παρατηρείται ότι, μέσα από τον δίπλευρο έλεγχο και την εναλλακτική υπόθεση $\theta \neq \theta_0$, καλύπτεται όλο το σύνολο των δυνατών τιμών της παραμέτρου θ.
- H_1: $\theta < \theta_0$, δηλαδή η παράμετρος είναι μικρότερη από την τιμή θ_0. Στην συγκεκριμένη περίπτωση πρόκειται για μονόπλευρο έλεγχο και συγκεκριμένα αριστερά μονόπλευρο έλεγχο. Στην προκειμένη περίπτωση, το σύνολο των δυνατών τιμών της παραμέτρου θ στη θ_0 περιορίζεται (H_1: $\theta < \theta_0$).
- H_1: $\theta > \theta_0$. Παρόμοια με τη προηγούμενη περίπτωση, πραγματοποιείται μονόπλευρο έλεγχος, ο οποίος αποκαλείται δεξιά μονόπλευρος έλεγχος στην παρούσα φάση και διατυπώνει ότι, οι τιμές της παραμέτρου θ είναι μεγαλύτερες της θ_0.

Στο σημείο αυτό αξίζει να σημειωθεί ότι, η επιλογή του μονόπλευρου ή του δίπλευρου ελέγχου εξαρτάται κάθε φορά από την έρευνα που επιθυμεί ο ενδιαφερόμενος να κάνει, καθώς επίσης και από το γεγονός, κατά πόσο είναι δυνατή η πρόβλεψη του αποτελέσματος.

Για παράδειγμα, αν θέλουμε να ελέγξουμε αν ο μέσος χρόνος εξυπηρέτησης των πελατών σε μία τράπεζα από ένα ταμία μ_1, υπό μελέτη μπορεί να φθάσει τον μέσο χρόνο εξυπηρέτησης των πελατών από έναν δεύτερο ταμία μ_2 (άριστος χρόνος, επομένως το μ_2 αποτελεί σημείο αναφοράς) τότε ο έλεγχος πρέπει να είναι μονόπλευρος (H_0: $\mu_1 \geq \mu_2$ και H_1: $\mu_1 < \mu_2$) για τον λόγο ότι δεν γίνεται $\mu_1 > \mu_2$, καθώς ο μέσος χρόνος εξυπηρέτησης των πελατών από τον πρώτο ταμία δεν μπορεί να ξεπεράσει τον αντίστοιχο χρόνο από το ταμία-σημείο αναφοράς.

Ωστόσο, στην περίπτωση που η σύγκριση που επιθυμείται να γίνει, αναφέρεται στον μέσο χρόνο μ_1, με τον μέσο χρόνο εξυπηρέτησης των πελατών σε κάποιο άλλο ταμία, έστω μ_3, ο έλεγχος δε θα πρέπει να είναι μονόπλευρος (H_0: $\mu_1 \leq \mu_3$ και H_1 : $\mu_1 > \mu_3$) αλλά δίπλευρος (H_0: $\mu_1 = \mu_3$ και H_1: $\mu_1 \neq \mu_3$) γιατί δεν γίνεται να αποκλειστεί το ενδεχόμενο $\mu_1 < \mu_3$, δηλαδή ο μέσος χρόνος εξυπηρέτησης στον πρώτο ταμία, να είναι μικρότερος από τον τρίτο ταμία.[2]

1.3 ΔΙΑΤΥΠΩΣΗ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ

Όπως ήδη αναφέρθηκε προηγουμένως, το πρώτο στάδιο για την ανάπτυξη των στατιστικών υποθέσεων είναι η διατύπωση μιας πρότασης - δήλωσης για κάποια συγκεκριμένη παράμετρο του πληθυσμού ή για περισσότερες αντίστοιχων πληθυσμών.

Οι στατιστικές υποθέσεις διατυπώνονται πάντοτε σε ζεύγη, προκειμένου το σύνολο του δειγματικού χώρου Ω να διαχωρίζεται σε δύο μικρότερα υποσύνολα Ω_1 και Ω_2. Για τα υποσύνολα αυτά ισχύει η εξής ιδιότητα:

$$\Omega = \Omega_1 \cap \Omega_2 \text{ και } \Omega_1 \cap \Omega_2 = \varnothing$$

Η παραπάνω ιδιότητα είναι ανεπτυγμένη με τέτοιο τρόπο, ώστε κατά την απόρριψη της μίας υπόθεσης να θεωρείται δεδομένη η αποδοχή της άλλης. Με τον τρόπο αυτό, η διενέργεια των στατιστικών υποθέσεων προσδίδει συνεπή αποτελέσματα από τον έλεγχο.

Στο σημείο αυτό σημειώνεται ότι, μια υπόθεση Η (H_0 ή H_1) θα αποκαλείται απλή, σε περίπτωση που το αντίστοιχο υποσύνολο Ω_0 ή Ω_1 αποτελείται από ένα μόνο σημείο. Σε διαφορετική περίπτωση, η υπόθεση ονομάζεται σύνθετη. Για παράδειγμα, αν η παράμετρος θ είναι η μέση τιμή (μ) μιας κατανομής και $\Omega = (-\infty, \infty)$, $\Omega_0=(0)$ και $\Omega_1 = (0,\infty)$, τότε η H_0 ($\mu=0$) είναι απλή υπόθεση, ενώ η H_1, ($\mu >0$) είναι σύνθετη υπόθεση.[3]

[2] Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β' Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη
http://www.aua.gr/gpapadopoulos/files/hypoth-tests-4.pdf

[3] http://users.auth.gr/dkugiu/Teach/CivilEngineer/hypothesis.pdf

Από τον παραπάνω έλεγχο η αποδοχή της μηδενικής υπόθεσης σημαίνει ότι το θ ανήκει στο υποσύνολο Ω_0.

Στην πρώτη περίπτωση, της αποδοχής της υπόθεσης, το εν λόγω υποσύνολο αποκαλείται περιοχή αποδοχής, ενώ το δεύτερο υποσύνολο αποτελεί την κρίσιμη περιοχή, κατά την οποία η υπόθεση H_0 απορρίπτεται και επομένως αποδεχόμαστε την υπόθεση H_1.

Για την αποσαφήνιση των παραπάνω εννοιών του ελέγχου υποθέσεων, χρησιμοποιείται το ακόλουθο παράδειγμα.

Μια εταιρεία γνωρίζει ότι η βελτίωση της σκληρότητας της επιφάνειας των εξαρτημάτων που παράγει, επιτυγχάνεται μόνο στο 25% των συνολικά παραγόμενων εξαρτημάτων από την διαδικασία που χρησιμοποιεί.

Επειδή το ποσοστό αυτό είναι μικρό η εταιρεία αποφάσισε να προβεί σε ένα πείραμα με μια διαφοροποιημένη διαδικασία, που θεωρείται πιο αποτελεσματική. Επιλέχθηκαν λοιπόν τυχαία 20 εξαρτήματα και υποβλήθηκαν στην νέα διαδικασία. Αν περισσότερα από 8 εξαρτήματα έχουν την επιθυμητή σκληρότητα μετά τη διαδικασία θα θεωρηθεί ότι η νέα διεργασία είναι ανώτερη απ’ αυτή που χρησιμοποιείται από την εταιρεία.

Η απαίτηση, ο αριθμός των βελτιωμένων εξαρτημάτων να είναι 8 είναι αυθαίρετος, αλλά επιλέγεται με σκοπό να εκφράζει ένα λογικό κέρδος ως προς τα 5 εξαρτήματα από τα 20, που πρόκειται να ήταν βελτιωμένα αν χρησιμοποιούταν η πρώτη διαδικασία.[4]

Η μηδενική υπόθεση H_0, είναι ότι η νέα διαδικασία είναι εξίσου αποτελεσματική με τη χρησιμοποιούμενη μέχρι τώρα. Η εναλλακτική υπόθεση είναι ότι η νέα διαδικασία είναι καλύτερη από την παλιά. Αυτό είναι ισοδύναμο με τον έλεγχο ότι η πιθανότητα επιτυχίας μιας δοκιμής είναι $p = 0,25$ έναντι της εναλλακτικής υπόθεσης ότι $p > 0,25$. Αυτό συνήθως γράφεται ως εξής:

H_0: $p = 0,25$

H_1: $p > 0,25$

[4] Ψώνιος Δ., (1999). «Στατιστική». Εκδόσεις Ζήτη, Θεσσαλονίκη
Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β’ Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη

Ο αριθμός των εξαρτημάτων διαιρείται σε δύο περιοχές (υποσύνολα). Στην μια περιοχή ανήκουν όλοι οι αριθμοί που είναι μικρότεροι ή ίσοι του 8 και στην άλλη περιοχή όσοι είναι μεγαλύτεροι του 8. Όλα τα δυνατά αποτελέσματα που είναι μεγαλύτερα του 8 αποτελούν την κρίσιμη περιοχή και όλα τα δυνατά αποτελέσματα που είναι μικρότερα ή ίσα του 8 ορίζουν την περιοχή αποδοχής.

Στην περίπτωση που το αποτέλεσμα της δοκιμής που πραγματοποιήθηκε με τη νέα διαδικασία είναι μεγαλύτερο του 8, απορρίπτουμε την υπόθεση H_0 και συνεπώς γίνεται αποδεκτή η εναλλακτική υπόθεση H_1. Σε περίπτωση που το αποτέλεσμα της δοκιμής είναι μικρότερο ή ίσο του 8, τότε αποδεχόμαστε την H_0 και απορρίπτεται η εναλλακτική υπόθεση H_1.[5]

1.4 ΠΕΡΙΟΧΗ ΑΠΟΡΡΙΨΗΣ ΚΑΙ ΚΡΙΣΙΜΟ ΣΗΜΕΙΟ

Όπως ήδη αναφέρθηκε σε προηγούμενη ενότητα, κάθε στατιστική υπόθεση διαχωρίζεται σε δύο μικρότερα υποσύνολα του δειγματικού χώρου Ω. Οι περιοχές αυτές αποτελούν την ***περιοχή αποδοχής*** (acceptance region) και την ***περιοχή απόρριψης*** (rejection region) της μηδενικής υπόθεσης.

Στο σημείο αυτό σημειώνεται ότι, η τιμή της παραμέτρου που διαχωρίζει την περιοχή αποδοχής από τη περιοχή απόρριψης ονομάζεται κρίσιμο σημείο (critical point) και συμβολίζεται με c.[6]

Κατά την διαδικασία του ελέγχου των υποθέσεων που έχουν δημιουργηθεί, το πρώτο βήμα είναι να θεωρηθεί ότι η μηδενική υπόθεση H_0 είναι σωστή. Με βάση την μηδενική υπόθεση H_0 προσδιορίζεται η στατιστική ελέγχου Q και η κατανομή της.

Στη περίπτωση που ο έλεγχος είναι παραμετρικός η κατανομή της Q προκύπτει από μετασχηματισμό της παραμέτρου θ. Η υπόθεση που γίνεται στην περίπτωση αυτή είναι, ότι υπάρχει κάποια κατανομή για την παράμετρο θ και η Q σχετίζεται άμεσα με τη θ. Στην περίπτωση μη-παραμετρικού ελέγχου, δε ισχύει η υπόθεση ύπαρξης κάποιας

[5] Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β' Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη

[6] http://www.stat-athens.aueb.gr/~jpan/statistiki-skepsi-II/chapter21.pdf

κατανομής για τη παράμετρο θ, επομένως, η κατανομή της Q βασίζεται σε άλλες ιδιότητες της θ.

Η κατανομή της στατιστικής ελέγχου Q δίνει την πιθανότητα η Q να πάρει κάποια τιμή (αν η q είναι διακριτή τυχαία μεταβλητή) ή να βρίσκεται σε ένα διάστημα τιμών, όταν η H_0 είναι αληθής. Το ενδεχόμενο η κατανομή να λάβει μία συγκεκριμένη τιμή, ή να βρίσκεται σε ένα διάστημα τιμών, εξαρτάται από το γεγονός εάν είναι διακριτή ή συνεχής τυχαία μεταβλητή.

Με βάση τις τιμές της κατανομής Q, μπορεί να εξαχθεί το εξής συμπέρασμα: όταν παρατηρούνται τιμές της Q που αντιστοιχούν σε μεγάλες πιθανότητες αυτό σημαίνει πως η H_0 είναι αληθής, ενώ αν οι τιμές της Q αντιστοιχούν σε μικρές πιθανότητες αυτό υποδηλώνει αμφιβολία για την ισχύ της μηδενικής υπόθεσης. Επομένως, συμπεραίνεται ότι, μη πιθανές τιμές της Q συνιστούν την απόρριψη της H_0.

Κατά κανόνα η περιοχή απόρριψης σχηματίζεται από τα άκρα της κατανομής της στατιστικής ελέγχου Q όπως αυτά ορίζονται από τις κρίσιμες τιμές. Αν ο έλεγχος είναι δίπλευρος τότε οι κρίσιμες τιμές $q_{\alpha/2}$ και $q_{1-\alpha/2}$ ορίζουν την περιοχή απόρριψης της H_0 ως εξής:

$$\boldsymbol{R = \{ q \mid q < q_{\alpha/2} \quad q > q_{1-\alpha/2} \},}$$

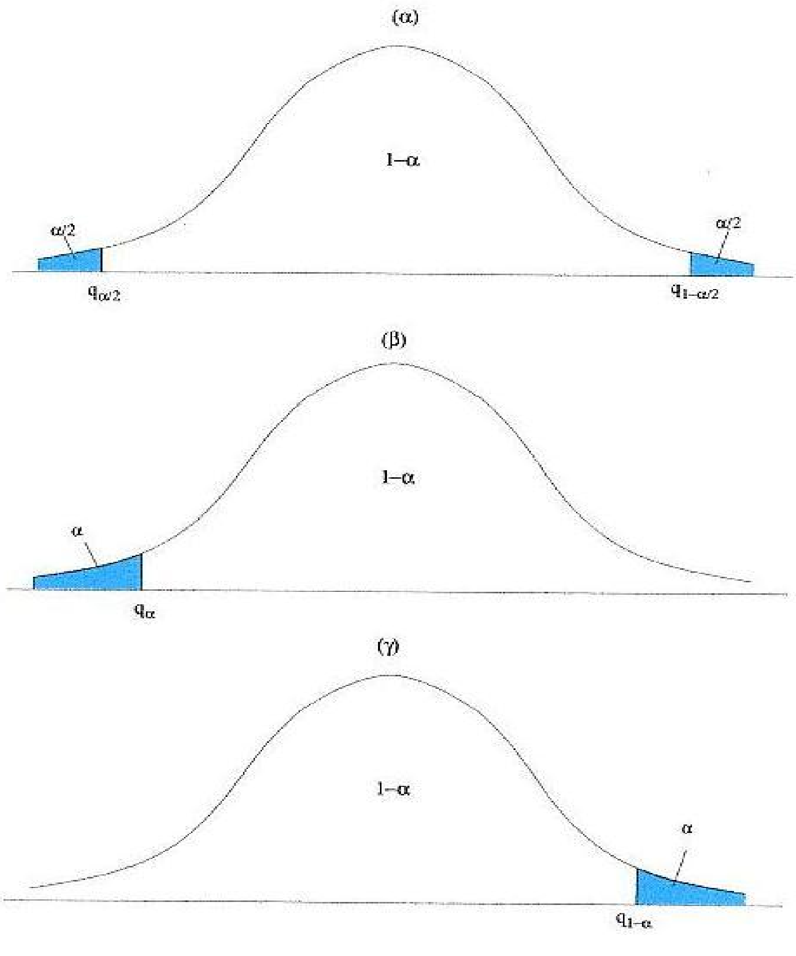
Η παραπάνω σχέση σχηματίζεται από τις δύο ουρές της κατανομής της Q με συνολική πιθανότητα α. Αν ο έλεγχος είναι μονόπλευρος τότε υπάρχει μόνο μία κρίσιμη τιμή, q_α ή $q_{1-\alpha}$, που ορίζει την περιοχή απόρριψης της H_0, ως εξής:

$\boldsymbol{R = \{ q \mid q < q_\alpha \}}$, για την αριστερή πλευρά

και

$\boldsymbol{R = \{ q \mid q > q_{1-\alpha} \}}$, για την δεξιά πλευρά.

Στο ακόλουθο σχήμα δίνονται οι περιοχές αποδοχής και απόρριψης για τον δίπλευρο και μονόπλευρο έλεγχο. Στο πρώτο διάγραμμα ο έλεγχος που πραγματοποιείται είναι δίπλευρος, ενώ στα δύο επόμενα σχήματα, πρόκειται για μονόπλευρο έλεγχο.[7]



Σχήμα 1.1

[7] Ρούσσας Γ. (1994). «Στατιστική συμπερασματολογία Τ. ΙΙ, Έλεγχος υποθέσεων». Εκδόσεις Ζήτη, Αθήνα

1.5 ΕΙΔΗ ΣΦΑΛΜΑΤΩΝ

Από την στιγμή που με βάση τα στατιστικά δεδομένα, αποφασίζεται η αποδοχή μιας υπόθεσης, υπάρχει η πιθανότητα, η απόφαση αυτή να είναι σωστή ή λανθασμένη. Η πιθανότητα λάθους αναφέρεται σε δύο τύπους, οι οποίοι είναι οι εξής:

Σφάλμα τύπου Ι. Σφάλμα τύπου Ι κάνουμε όταν έχουμε απορρίψει την μηδενική υπόθεση, ενώ στην πραγματικότητα αυτή είναι αληθής. Κατά την διάρκεια ενός στατιστικού ελέγχου, η μέγιστη πιθανότητα με την οποία δεχόμαστε να κάνουμε σφάλμα τύπου Ι, ονομάζεται επίπεδο σημαντικότητας και συμβολίζεται με α.

Σφάλμα τύπου ΙΙ. Σφάλμα τύπου ΙΙ, προκύπτει κάθε φορά που αποδεχόμαστε (ή πιο σωστά, δεν απορρίπτουμε) την μηδενική υπόθεση, ενώ στην πραγματικότητα αυτή είναι εσφαλμένη. Όπως συμβαίνει και στο σφάλμα τύπου Ι, έτσι και εδώ, η πιθανότητα μια εσφαλμένη υπόθεση H_0 να γίνει αποδεκτή συμβολίζεται με β.

Καθίσταται λοιπόν σαφές, ότι και στις δύο παραπάνω περιπτώσεις επιλέγεται η λανθασμένη απόφαση. Τα ανωτέρω ενδεχόμενα παρουσιάζονται στον πίνακα που ακολουθεί.[8]

Πίνακας 1.1

Στατιστικός έλεγχος	Πραγματική κατάσταση	
	Η H_0 είναι ορθή	Η H_0 είναι εσφαλμένη
Η H_0 γίνεται αποδεκτή	*Ορθή απόφαση* *Πιθανότητα 1-α*	***Σφάλμα τύπου ΙΙ*** ***Πιθανότητα β***
Η H_0 απορρίπτεται	***Σφάλμα τύπου Ι*** ***Πιθανότητα α***	*Ορθή απόφαση* *Πιθανότητα 1-β*

[8] http://users.auth.gr/dkugiu/Teach/CivilEngineer/hypothesis.pdf

Από τα παραπάνω μπορεί να εξαχθεί το συμπέρασμα ότι, κατά την διαδικασία του ελέγχου υποθέσεων, την καλύτερη δυνατή απόφαση θα μπορούσε να λάβει κάποιος, αν μπορούσε να μηδενίσει την πιθανότητα απόρριψης της H_0 ενώ αυτή είναι ορθή, καθώς επίσης, να μηδενίσει την πιθανότητα αποδοχής της H_0 ενώ αυτή είναι εσφαλμένη.

Με άλλα λόγια, σε κάθε έλεγχο στατιστικών υποθέσεων θα πρέπει να επιδιώκεται η ελαχιστοποίηση των πιθανοτήτων α και β. Ωστόσο, για να γίνει κάτι τέτοιο θα πρέπει το δείγμα να αποτελείται από ολόκληρο τον πληθυσμό και όχι από ένα μέρος του.

Επομένως, προκειμένου μια διαδικασία ελέγχου υποθέσεων να είναι ικανοποιητική, θα πρέπει να τείνει στην ελαχιστοποίηση των παραπάνω σφαλμάτων. Στο σημείο αυτό αξίζει να σημειωθεί ότι, κάτι τέτοιο δεν αποτελεί και εύκολη υπόθεση, για τον λόγο ότι, συνήθως η προσπάθεια για μείωση της πιθανότητας σφάλματος σε μία διαδικασία ελέγχου, ενδέχεται να αυξήσει την πιθανότητα σφάλματος του άλλου τύπου. Προκειμένου λοιπόν να μην δημιουργούνται τέτοιες καταστάσεις, επιδιώκεται ένα είδος χρυσής τομής και κυρίως, η μείωση του περισσότερο σοβαρού σφάλματος.

Όπως ήδη αναφέρθηκε, ο μόνος τρόπος να μειώσουμε και τα δύο είδη των σφαλμάτων είναι να αυξήσουμε το μέγεθος του δείγματος, γεγονός που δεν είναι τόσο εφικτό στην πραγματικότητα. Συνεπώς, παρόλο που επιδιώκεται η διατήρηση του σφάλματος τύπου Ι και ΙΙ σε ένα ελάχιστο επίπεδο και για ορισμένο μέγεθος δείγματος n, δεν είναι εύκολο να ελεγχθούν και τα δύο ταυτόχρονα.[9]

Σημειώνεται επίσης, ότι επειδή η αποφυγή του σφάλματος τύπου Ι είναι πιο σημαντική από αυτήν του τύπου ΙΙ, λόγω του ότι απορρίπτεται μια ορθή υπόθεση, το κριτήριο που χρησιμοποιείται θα πρέπει να είναι εκείνο, το οποίο σπανίως έχει ως αποτέλεσμα την απόρριψη μιας αληθινής υπόθεσης.

Στην περίπτωση που η σύγκριση δύο κριτηρίων έχει την ίδια πιθανότητα α για να διενεργηθεί σφάλμα τύπου Ι, τότε επιλέγεται εκείνο, για το οποίο η πιθανότητα διεξαγωγής σφάλματος τύπου ΙΙ θα είναι πιο μικρή.

Συμπερασματικά, κατά την διαδικασία ελέγχου των υποθέσεων, θα πρέπει να γίνει κατανοητό ότι η αποδοχή μιας υπόθεσης απλώς σημαίνει ότι τα στοιχεία του δείγματος δεν παρέχουν επαρκείς ενδείξεις για να την απορρίψουμε. Εξάλλου, η

[9] Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β’ Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη

απόρριψη της υπόθεσης, σημαίνει απλά ότι τα στοιχεία του δείγματος μας παρέχουν ενδείξεις ότι δεν μπορούμε να τη δεχθούμε.[10]

1.6 ΙΣΧΥΣ ΤΟΥ ΕΛΕΓΧΟΥ

Η πιθανότητα απόρριψης της μηδενικής υπόθεσης H_0, όταν αυτή είναι εσφαλμένη, γιατί η H_1 είναι αληθής ονομάζεται ***ισχύς του ελέγχου.*** Η ισχύς του ελέγχου ισούται με 1-β και προέρχεται από την πιθανότητα απόρριψης της εσφαλμένης υπόθεσης H_0, όπως φαίνεται και στον πίνακα 1.1. Καθίσταται σαφές ότι όσο μεγαλύτερη τιμή λαμβάνει η ισχύς του ελέγχου, τόσο καλύτερη θα είναι η απόφαση που θα ληφθεί.

Όπως έχει προαναφερθεί, η πιθανότητα 1-β εκφράζει την πιθανότητα λήψης της ορθής απόφασης, όταν η H_1 είναι αληθής. Στο σημείο αυτό σημειώνεται ότι, η πιθανότητα 1-β μεταβάλλεται καθώς μεταβάλλεται η τιμή της παραμέτρου που ελέγχεται με την υπόθεση H_1. Αυτό συμβαίνει, για τον εξής λόγο: όταν η τιμή της παραμέτρου που ελέγχεται με την υπόθεση H_1 είναι απομακρυσμένη από την τιμή της παραμέτρου σύμφωνα με την υπόθεση H_0, η πιθανότητα αποδοχής της εσφαλμένης απόφασης, δηλαδή η πιθανότητα β, θα είναι μικρή. Επομένως, το 1-β που εκφράζει τη πιθανότητα να κάνουμε αποδεκτή την H_1, όταν είναι ορθή, θα είναι μεγάλο. Γίνεται σαφές λοιπόν ότι σε αντίθετη περίπτωση το 1-β θα είναι μικρό.

Συνεπώς, οι τιμές των πιθανοτήτων β και 1-β είναι συναρτήσεις των τιμών της παραμέτρου που ελέγχονται κάτω από την εναλλακτική υπόθεση. Για τον λόγο αυτό, στη βιβλιογραφία το 1-β αποκαλείται και ως συνάρτηση ισχύος του ελέγχου.[11]

---

[10] Ψώνιος Δ., (1999). «Στατιστική». Εκδόσεις Ζήτη, Θεσσαλονίκη
http://www.aua.gr/gpapadopoulos/files/hypoth-tests-4.pdf
[11] http://www.stat-athens.aueb.gr/~jpan/statistiki-skepsi-II/chapter21.pdf

1.7 ΠΑΡΑΤΗΡΟΥΜΕΝΟ ΕΠΙΠΕΔΟ ΣΗΜΑΝΤΙΚΟΤΗΤΑΣ- Η ΤΙΜΗ P-VALUE

Από την παραπάνω ανάλυση εξάγεται το συμπέρασμα, ότι δεν έχει διαπιστωθεί κάποιος κανόνας, ο οποίος να οδηγεί στην επιλογή του κατάλληλου επιπέδου σημαντικότητας κατά την διαδικασία των στατιστικών ελέγχων.

Για τον λόγο αυτό οι στατιστικοί επιστήμονες έχουν οριστεί ένα μέγεθος για την καλύτερη περιγραφή της κατάστασης. Το μέγεθος αυτό είναι το παρατηρούμενο επίπεδο σημαντικότητας, ή αλλιώς τιμή p-value.

Ως παρατηρούμενο επίπεδο σημαντικότητας (p-value), ορίζεται η πιθανότητα, η στατιστική συνάρτηση ελέγχου να λάβει μία τιμή τόσο ακραία, ή περισσότερο ακραία από αυτήν που πήρε για το συγκεκριμένο δείγμα, κάτω από την μηδενική υπόθεση.

Ουσιαστικά, η τιμή p-value χρησιμοποιείται αντί του επιπέδου σημαντικότητας α, χωρίς ωστόσο να υπάρχει κάποια μεταβολή στην θεωρία ανάπτυξης της διαδικασίας ελέγχου των υποθέσεων. Με την χρησιμοποίηση της τιμής p-value, ο ερευνητής καθιστά στον ενδιαφερόμενο γνωστό το επίπεδο σημαντικότητας και αφήνει σε αυτόν την επιλογή κατά πόσο θα πρέπει να απορριφθεί ή όχι η μηδενική υπόθεση. Με άλλα λόγια, η τιμή p-value, αναφέρεται σε ένα επίπεδο σημαντικότητας, το οποίο πλέον επιλέγεται από τον ενδιαφερόμενο και όχι τον ερευνητή.

Στο σημείο αυτό αξίζει να σημειωθεί ότι, η τιμή p-value, δεν αντικατοπτρίζει την πιθανότητα η μηδενική υπόθεση να είναι ορθή. [12]

1.8 ΔΙΑΔΙΚΑΣΙΑ ΣΤΑΤΙΣΤΙΚΟΥ ΕΛΕΓΧΟΥ ΚΑΙ ΕΛΕΓΧΟΥ ΣΗΜΑΝΤΙΚΟΤΗΤΑΣ

Η διαδικασία του στατιστικού ελέγχου περιλαμβάνει κάποια στάδια τα οποία συνοψίζονται στην συνέχεια της ενότητας.

[12] Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β’ Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη

Στο πρώτο στάδιο, ορίζεται η μηδενική υπόθεση H_0, καθώς και η εναλλακτική της H_1. Όπως έχει ήδη αναφερθεί, η διατύπωση της μηδενικής υπόθεσης έχει την μορφή H_0: $\theta=\theta_0$ και η εναλλακτική υπόθεση H_1, εξαρτάται από τον τύπο του ελέγχου που επιθυμείται. Υπενθυμίζεται ότι, οι μορφές ελέγχου των υποθέσεων είναι ο δίπλευρος έλεγχος και ο μονόπλευρος έλεγχος. Δηλαδή, η εναλλακτική υπόθεση μπορεί να έχει την εξής μορφή: H_1: $\theta > \theta_0$ ή H_1: $\theta < \theta_0$, ή H_1: $\theta \neq \theta_0$.

Στην συνέχεια επιλέγεται το επίπεδο σημαντικότητας α, καθώς και η κατάλληλη στατιστική ελέγχου Q. Ακολούθως, υπολογίζεται η κρίσιμη τιμή της Q από αντίστοιχο στατιστικό πίνακα και καθορίζεται η περιοχής απόρριψης R. Στο σημείο αυτό, επιλέγεται ένα δείγμα και υπολογίζεται μία τιμή της εκτιμήτριας $\widehat{\Theta}$, της παραμέτρου θ. κατόπιν, επιλέγεται το επίπεδο σημαντικότητας α και καθορίζεται η κρίσιμη περιοχή από την σχέση:

$$P \{ \bar{\theta} \text{ €} \, \Omega \mid H_0 \} = \alpha.$$

Η παραπάνω σχέση δηλώνει ότι η πιθανότητα να ανήκει η $\hat{\theta}$ στο σύνολο Ω δεδομένου ότι η H_0 είναι αληθινή ισούται με την πιθανότητα α.

Αναφορικά με την κρίσιμη περιοχή και το επίπεδο σημαντικότητας θα πρέπει να σημειώσουμε τα εξής: στην περίπτωση που η εναλλακτική υπόθεση H_1 είναι μονόπλευρη, δηλαδή H_1: $\theta > \theta_0$, τότε και η κρίσιμη περιοχή είναι επίσης μονόπλευρη. Επομένως, σε περίπτωση στατιστικού ελέγχου μίας δεξιάς μονόπλευρης εναλλακτικής υπόθεσης, η κρίσιμη περιοχή ορίζεται από την ακόλουθη σχέση:

$$P \{ \hat{\theta} - \theta_0 > d(\alpha) \} = \alpha.$$

Στην περίπτωση που η εναλλακτική υπόθεση H_1 είναι δίπλευρη, δηλαδή η H_1 έχει διατυπωθεί ως $H_1 : \theta \neq \theta_0$, τότε η κρίσιμη περιοχή είναι επίσης δίπλευρη και υπολογίζεται από την σχέση:

$$P \{ |\hat{\theta} - \theta_0| > d(\tfrac{1}{2}\alpha) \} = \alpha,$$

όπου, d είναι οι τιμές που καθορίζονται από το α και το είδος της κατανομής της εκτιμήτριας $\hat{\Theta}$.

Στο τελικό στάδιο και εφ' όσον έχει υπολογιστεί η στατιστική ελέγχου, προβαίνουμε σε απόρριψη της H_0 αν η q ανήκει στην περιοχή απόρριψης R, ή αποδοχή της H_0 αν η q δεν ανήκει στην περιοχή απόρριψης R.

Η απόρριψη της μηδενικής υπόθεσης H_0 σημαίνει ότι, η διαφορά ανάμεσα στην υποθετική τιμή θ_0 και την τιμή της εκτιμήτριας $\hat{\Theta}$ είναι στατιστικά σημαντική. Αντίθετα, η αποδοχή της μηδενικής υπόθεσης H_0 σημαίνει ότι η διαφορά ανάμεσα στην τιμή της εκτιμήτριας $\hat{\Theta}$ και στη θ_0 δεν είναι στατιστικά σημαντική.[13]

[13] http://users.auth.gr/dkugiu/Teach/CivilEngineer/hypothesis.pdf

1.9 ΔΙΑΣΤΗΜΑΤΑ ΕΜΠΙΣΤΟΣΥΝΗΣ ΚΑΙ ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ

Σύμφωνα με την θεωρία των διαστημάτων εμπιστοσύνης, εκτιμάται μία παράμετρος του πληθυσμού, για παράδειγμα, η μέση τιμή και στην συνέχεια δημιουργείται ένα διάστημα εμπιστοσύνης που θα περιέχει μέσα την εν λόγω εκτίμηση. Κατά την δημιουργία των διαστημάτων εμπιστοσύνης δεν λαμβάνονται υπόψη καθόλου πληροφορίες σχετικά με τον πληθυσμό. Το διάστημα εμπιστοσύνης παρέχει βεβαιότητα ότι, κατά ένα ποσοστό η παράμετρος που επιδιώκεται βρίσκεται μέσα σε ένα εύρος τιμών.

Όταν υπάρχουν πληροφορίες σχετικά με την παράμετρο που έχει χρησιμοποιηθεί, τότε δίνεται η δυνατότητα δημιουργίας υποθέσεων για την παράμετρο. Στην περίπτωση αυτή πρόκειται για το δεύτερο βασικό εργαλείο της στατιστικής συμπερασματολογίας, τον έλεγχο των υποθέσεων. Όπως ήδη είναι γνωστό, σκοπός της ανάπτυξης των στατιστικών ελέγχων είναι η αποδοχή ή η απόρριψη μίας υπόθεσης, χρησιμοποιώντας τις πληροφορίες που παρέχονται για το δείγμα.[14]

[14] Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β’ Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη

ΚΕΦΑΛΑΙΟ ΔΕΥΤΕΡΟ
ΕΛΕΓΧΟΣ ΣΤΑΤΙΣΤΙΚΩΝ ΥΠΟΘΕΣΕΩΝ

2.1 ΕΙΣΑΓΩΓΗ

Στο κεφάλαιο που ακολουθεί, αναλύεται ο έλεγχος των στατιστικών υποθέσεων για διάφορες παραμέτρους. Αρχικά, πραγματοποιούνται έλεγχοι υποθέσεων για τη μέση τιμή, τόσο με γνωστή, όσο και άγνωστη τυπική απόκλιση.

Έλεγχος πραγματοποιείται επίσης για τη διασπορά, την αναλογία, τη διαφορά των μέσων τιμών, των διασπορών, καθώς επίσης και των αναλογιών.

2.2 ΕΛΕΓΧΟΣ ΠΑΡΑΜΕΤΡΩΝ ΓΙΑ ΕΝΑ ΠΛΗΘΥΣΜΟ
2.2.1 ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ ΓΙΑ ΤΗΝ ΜΕΣΗ ΤΙΜΗ

Στην συνέχεια, περιγράφεται ο έλεγχος υποθέσεων με γνωστή και άγνωστη τυπική απόκλιση.

Γνωστή τυπική απόκλιση

Στην συγκεκριμένη περίπτωση ο έλεγχος υποθέσεων πραγματοποιείται μέσω της τυπικής απόκλισης σ. Έστω λοιπόν, ότι για έναν πληθυσμό δεν γνωρίζουμε την μέση τιμή του, αλλά την τυπική του απόκλιση. Επιπλέον, γίνεται η υπόθεση ότι η μέση τιμή για τον πληθυσμό είναι μ_o.

Στο σημείο αυτό θα διενεργηθεί έλεγχος, κατά πόσο η τιμή αυτή θεωρείται μπορεί να γίνει αποδεκτή.

Η μηδενική υπόθεση που έχει δημιουργηθεί είναι η εξής:

$$H_0: \mu = \mu_0$$

Σε προηγούμενο κεφάλαιο αναφέρθηκε ότι, η εναλλακτική υπόθεση H_1 μπορεί να λάβει μία από τις ακόλουθες μορφές:

i. H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$ (δίπλευρος έλεγχος)

ii. H_0: $\mu = \mu_0$ H_1: $\mu > \mu_0$ (μονόπλευρος προς τα πάνω)

iii. H_0: $\mu = \mu_0$ H_1: $\mu < \mu_0$ (μονόπλευρος προς τα κάτω)

Από τα παραπάνω συμπεραίνεται ότι, η μέση τιμή ενός πληθυσμού μ_0 μπορεί να λάβει ως εναλλακτικές υποθέσεις: α) η μέση τιμή του πληθυσμού να είναι διάφορη μ_0, β) μεγαλύτερη του μ_0, γ) μικρότερη του μ_0. Καθίσταται λοιπόν σαφές ότι οι πιθανοί έλεγχοι υποθέσεων που μπορούν να πραγματοποιηθούν είναι τρεις.

Αφού, προσδιοριστεί η εναλλακτική υπόθεση, στην συνέχεια λαμβάνεται ένα δείγμα από τον πληθυσμό και κατόπιν υπολογίζεται η μέση τιμή του. Έστω λοιπόν, ότι η μέση τιμή, συμβολίζεται με $\bar{x}$.

Στο σημείο αυτό σημειώνεται ότι σύμφωνα με το Κεντρικό Οριακό Θεώρημα, το άθροισμα καθώς και η μέση τιμή, μεγάλου αριθμού ανεξάρτητων παρατηρήσεων, ακολουθεί κατά προσέγγιση την κανονική κατανομή, ανεξάρτητα από το ποια κατανομή ακολουθούν οι παρατηρήσεις.

Επομένως, η κατανομή της εκτιμήτριας $\bar{X}$ είναι κανονική και η μεταβλητή

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

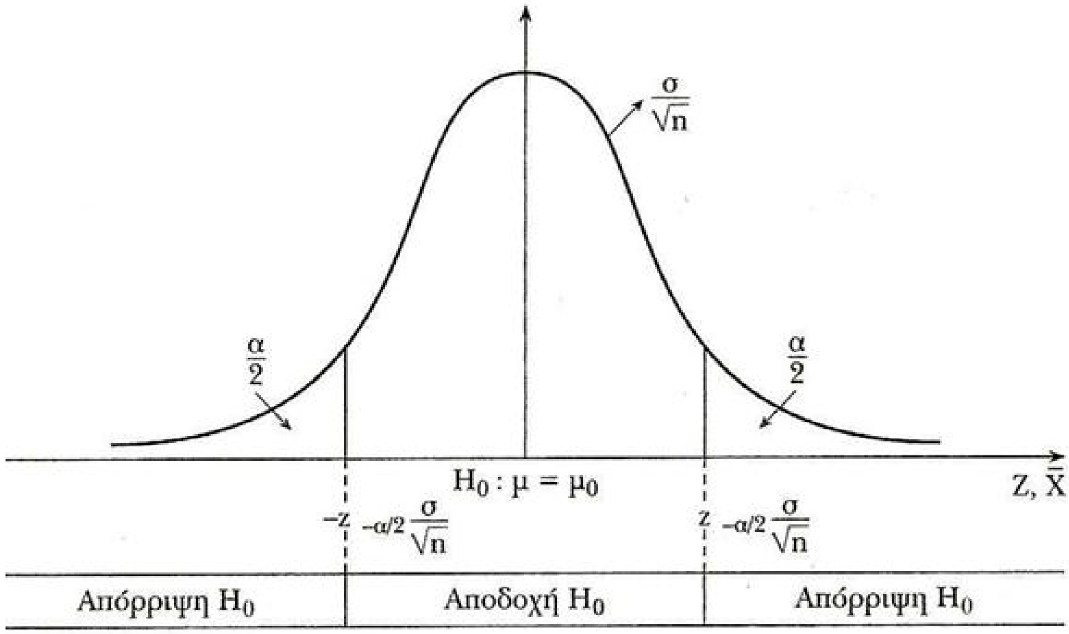
είναι η τυπική κανονική μεταβλητή, δηλαδή η Z κατανέμεται ως N (0,1).

Στην συνέχεια καθορίζεται το επίπεδο σημαντικότητας α, για τον έλεγχο της υπόθεσης. Η εναλλακτική υπόθεση που ελέγχεται είναι, η H_1: $\mu \neq \mu_0$, που σημαίνει ότι, πολύ μικρές ή πολύ μεγάλες τιμές της $\bar{X}$ και συνεπώς και της $\bar{X}-\mu_0$, οδηγούν στην απόρριψη της H_0 και στην αποδοχή της H_1. Η περιοχή απόρριψης, για επίπεδο σημαντικότητας α, ορίζεται από τις σχέσεις:

$$\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > Z_{-\alpha/2} \quad \text{και} \quad \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < Z_{\alpha/2} \quad \text{ή } |Z| \geq Z_{\alpha/2} \qquad (2.1)$$

Στο σχήμα 2.1 που ακολουθεί παρουσιάζεται διαγραμματικά παραπάνω έλεγχος, με εναλλακτική υπόθεση την H_1: $\mu \neq \mu_0$.

Σημειώνεται ότι, για τους ελέγχους των υπόλοιπων εναλλακτικών υποθέσεων, η διαδικασία είναι αντίστοιχη με αυτή που περιγράφηκε παραπάνω.



Σχήμα 2.1

Στην περίπτωση ο έλεγχος υπόθεσης είναι ο έλεγχος (ii), κατά τον οποίο η μηδενική υπόθεση είναι H_0: $\mu = \mu_0$ και η εναλλακτική υπόθεση H_1: $\mu > \mu_0$, τότε η απόρριψη της μηδενικής υπόθεσης γίνεται μόνο για πολύ μεγάλες τιμές της Z, εξαιτίας μεγάλης τιμής της $\bar{X}$ και συνεπώς και της $\bar{X}-\mu_0$.

Δηλαδή, η απόρριψη της μηδενικής υπόθεσης και η αποδοχή της εναλλακτικής υπόθεσης, γίνεται μόνο όταν ισχύει η παρακάτω ανισότητα:[15]

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > Z_\alpha \qquad (2.2)$$

[15] Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β' Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη
http://users.auth.gr/dkugiu/Teach/CivilEngineer/hypothesis.pdf

Τέλος, στην περίπτωση που ελέγχουμε την δεύτερη μονόπλευρη υπόθεση (iii), κατά την οποία H_0: $\mu = \mu_0$ και H_1: $\mu < \mu_0$, η απόρριψη της H_0 και αποδοχή της H_1, πραγματοποιείται μόνο για πολύ μικρές τιμές της $\bar{X}$ και συνεπώς της $\bar{X}$-μ_0 και της Z.

Η ανισότητα που πρέπει να ικανοποιείται στην περίπτωση αυτή είναι η εξής:[16]

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} < -Z_\alpha \qquad (2.3)$$

Στο σημείο αυτό σημειώνεται ότι, στην κανονική κατανομή, η μέση τιμή, η διάμεσος και η επικρατούσα τιμή ισούται, καθώς πρόκειται για μία συμμετρική κατανομή. Σημειώνεται λοιπόν, ότι όλες οι τιμές των παραμέτρων της κεντρικής τάσης συμπίπτουν.

Άγνωστη τυπική απόκλιση

Πρόκειται για την δεύτερη περίπτωση ελέγχου υποθέσεων για την μέση τιμή ενός πληθυσμού. Μέχρι τώρα η τυπική απόκλιση ήταν γνωστή. Στην παρούσα ενότητα θα πραγματοποιηθεί έλεγχος υποθέσεων για πληθυσμούς των οποίων η τυπική απόκλιση είναι άγνωστη.

Στην περίπτωση που το δείγμα που εξετάζεται είναι μεγάλο, δηλαδή, αν $n>30$, τότε για την διενέργεια του έλεγχου, αντικαθιστούμε απλά την μεταβλητή σ με την εκτίμηση της τυπικής απόκλισης του δείγματος, s, στις σχέσεις (2.1), (2.2) και (2.3) που περιγράφηκαν παραπάνω. Σημειώνεται λοιπόν, ότι στην περίπτωση που η τυπική απόκλιση είναι άγνωστη αλλά το δείγμα είναι μεγάλο, ακολουθείται η κανονική κατανομή.

Σε περίπτωση όμως, που το δείγμα είναι μικρό ($n<30$), τότε ο έλεγχος της υπόθεσης για την μέση τιμή θα πραγματοποιηθεί με τη βοήθεια της μεταβλητής t, η οποία ακολουθεί την κατανομή του Student και ορίζεται από την εξής σχέση:

---

[16] Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β' Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη
http://users.auth.gr/dkugiu/Teach/CivilEngineer/hypothesis.pdf

$$t = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \sim t_{n-1} \qquad (2.4)$$

Στην περίπτωση που επιλέγεται να χρησιμοποιηθεί η κατανομή του Student στον έλεγχο των υποθέσεων, οι αντίστοιχες σχέσεις των παραπάνω (2.1), (2.2), και (2.3) είναι οι εξής:

$$\frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} < t_{\nu,\alpha/2} \quad \text{και} \quad \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} < - t_{\nu,\alpha/2} \qquad (2.5)$$

$$\frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} > + t_{\nu,\alpha} \qquad (2.6)$$

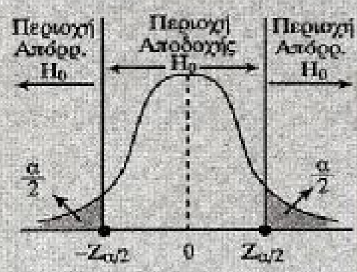
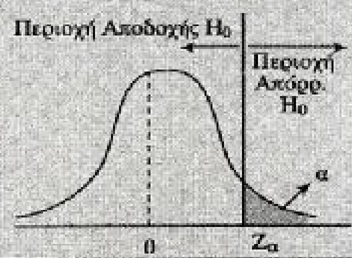
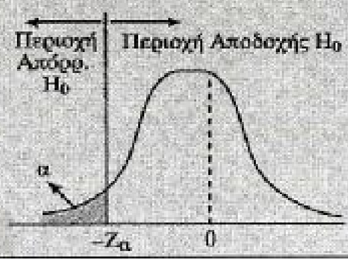
$$\frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} > - t_{\nu,\alpha} \qquad (2.7)$$

όπου $\nu = n-1$ είναι οι βαθμοί ελευθερίας. Στο σημείο αυτό αξίζει να σημειωθεί ότι, οι βαθμοί ελευθερίας δηλώνουν τις τυχαίες μεταβλητές για το πρόβλημα που εξετάζεται.

Οι παραπάνω σχέσεις (2.5), (2.6) και (2.7), καθορίζουν τις κρίσιμες περιοχές. Δηλαδή, όταν οι εν λόγω σχέσεις ικανοποιούνται σε επίπεδο σημαντικότητας α, τότε η μηδενική υπόθεση H_0 απορρίπτεται και αποδεχόμαστε την εναλλακτική υπόθεση H_1.

Στους δύο πίνακες που ακολουθούν αναγράφεται το είδος του έλεγχου, τα κριτήρια ελέγχου, καθώς και η περιοχή αποδοχής και απόρριψης της μηδενικής υπόθεσης, τόσο για την περίπτωση που το δείγμα είναι μεγάλο (πίνακας 2.1 - κανονική κατανομή), όσο και για την περίπτωση που το δείγμα είναι μικρό (πίνακας 2.2 - κατανομή t-student).

Πίνακας 2.1 Κριτήρια ελέγχου ανάλογα της μορφής της εναλλακτικής

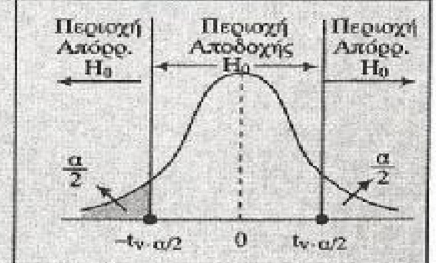
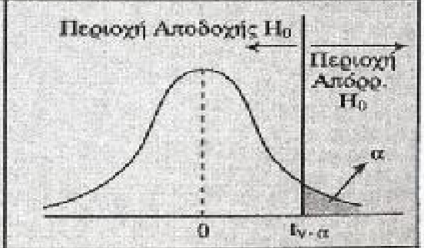
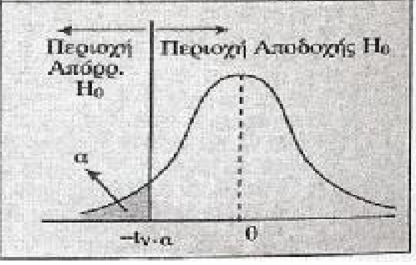
Είδος Ελέγχου	**Κριτήριο Ελέγχου:** Απορρίπτεται η H_0 εάν:	**Περιοχή Αποδοχής & Απόρριψης της βασικής υπόθεσης (H_0)**
α) **Δίπλευρος:** H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$	$\lvert Z^* \rvert > \lvert z_{\alpha/2} \rvert$ ή είτε $Z^* > z_{\alpha/2}$ είτε $Z^* < -z_{\alpha/2}$	
β) **Μονόπλευρος Πάνω:** H_0: $\mu = \mu_0$ H_1: $\mu > \mu_0$	$Z^* > z_\alpha$	
γ) **Μονόπλευρος Κάτω:** H_0: $\mu = \mu_0$ H_1: $\mu < \mu_0$	$Z^* < -z_\alpha$	

Πηγή: Ρούσσας Γ. (1994). «Στατιστική συμπερασματολογία Τ. ΙΙ, Έλεγχος υποθέσεων». Εκδόσεις Ζήτη, Αθήνα

Τα κριτήρια ελέγχου ανάλογα με την μορφή της εναλλακτικής υπόθεσης, καθώς και η περιοχή αποδοχής και απόρριψης της μηδενικής υπόθεσης, με βάση την κατανομή t-student παρουσιάζονται στον πίνακα 2.2 που ακολουθεί.[17]

[17] Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β' Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη
http://users.auth.gr/dkugiu/Teach/CivilEngineer/hypothesis.pdf

Πίνακας 2.2 Κριτήρια ελέγχου ανάλογα της μορφής της εναλλακτικής

Είδος Ελέγχου	Κριτήριο Ελέγχου: Απορρίπτεται η H_0 όταν:	Περιοχή Αποδοχής & Απόρριψης της (H_0)
α) Δίπλευρος: H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$	$\lvert t_v^* \rvert > \lvert t_{v \cdot (\alpha/2)} \rvert$ ή είτε $t_v^* > t_{v \cdot (\alpha/2)}$ είτε $t_v^* < -t_{v \cdot (\alpha/2)}$	
β) Μονόπλευρος Πάνω: H_0: $\mu = \mu_0$ H_1: $\mu > \mu_0$	$t_v^* > t_{v \cdot \alpha}$	
γ) Μονόπλευρος Κάτω: H_0: $\mu = \mu_0$ H_1: $\mu < \mu_0$	$t_v^* < -t_{v \cdot \alpha}$	

Πηγή: Ρούσσας Γ. (1994). «Στατιστική συμπερασματολογία Τ. ΙΙ, Έλεγχος υποθέσεων». Εκδόσεις Ζήτη, Αθήνα

2.2.2 ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ ΓΙΑ ΤΗ ΔΙΑΣΠΟΡΑ

Στην συνέχεια του κεφαλαίου, διενεργούνται οι έλεγχοι υποθέσεων για τη διασπορά ενός πληθυσμού. Όπως ήδη αναφέρθηκε, οι πιθανοί έλεγχοι που μπορεί να προκύψουν από την διατύπωση της εναλλακτικής υποθέσεις είναι τρεις. Ακολούθως διατυπώνονται οι υποθέσεις για την διασπορά ενός πληθυσμού.

i) Δίπλευρος έλεγχος:

$H_0: \sigma^2 = \sigma_0^2$ ή για την τυπική απόκλιση $H_0: \sigma = \sigma_0$

$H_1: \sigma^2 \neq \sigma_0^2$ $H_1: \sigma \neq \sigma_0$

ii) Μονόπλευρος Πάνω :

$H_0: \sigma^2 = \sigma_0^2$ ή για την τυπική απόκλιση $H_0: \sigma = \sigma_0$

$H_1: \sigma^2 > \sigma_0^2$ $H_1: \sigma > \sigma_0$

iii) Μονόπλευρος Κάτω :

$H_0: \sigma^2 = \sigma_0^2$ ή για την τυπική απόκλιση $H_0: \sigma = \sigma_0$

$H_1: \sigma^2 < \sigma_0^2$ $H_1: \sigma < \sigma_0$

όπου σ_0^2 είναι η γνωστή πληθυσμιακή διασπορά.

Κάτω από την υπόθεση της κανονικότητας του πληθυσμού κι εφόσον το δείγμα είναι τυχαίο, η στατιστική ελέγχου της διασποράς υπολογίζεται από την $X_\nu^2 \sim X_\nu^2$ με $\nu = (n - 1)$.

Έτσι, για τον έλεγχο της υπόθεσης, η μεταβλητή που υπολογίζεται είναι η ακόλουθη:

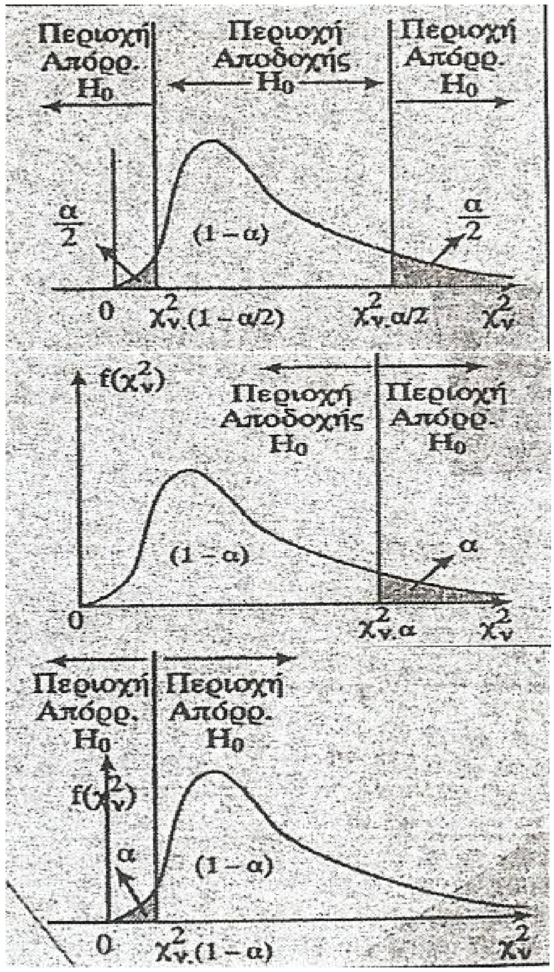
$$X_\nu^2 = \frac{(n-1) \cdot s^2}{\sigma_0^2}$$

Η αντίστοιχη στατιστική ελέγχου της τυπικής απόκλισης (σ) είναι παρόμοια με τον έλεγχο υπόθεσης για τη διασπορά (σ^2). Ουσιαστικά, στην συγκεκριμένη περίπτωση υπολογίζουμε την τετραγωνική ρίζα αυτού του αποτελέσματος, ως έλεγχο της τυπικής απόκλισης (σ).

Για την διασπορά, τα κριτήρια ελέγχου, ανάλογα τη μορφή του, δίνονται στον παρακάτω πίνακα σύμφωνα με την X^2. [18]

ΠΙΝΑΚΑΣ 2.3

Είδος Ελέγχου	**Κριτήριο Ελέγχου: Απορρίπτεται η H_0 όταν:**	**Περιοχές Αποδοχής & Απόρριψης της H_0**
α) Δίπλευρος: H_0: $\sigma^2 = \sigma_0^2$ H_1: $\sigma^2 \neq \sigma_0^2$	είτε $\chi^2_{v^*} > \chi^2_{v\cdot(\alpha/2)}$ είτε $\chi^2_{v^*} < \chi^2_{v\cdot(1-\alpha/2)}$	
β) Μονόπλευρος Πάνω: H_0: $\sigma^2 = \sigma_0^2$ H_1: $\sigma^2 > \sigma_0^2$	$\chi^2_{v^*} > \chi^2_{v\cdot\alpha}$	
γ) Μονόπλευρος Κάτω: H_0: $\sigma^2 = \sigma_0^2$ H_1: $\sigma^2 < \sigma_0^2$	$\chi^2_{v^*} < \chi^2_{v\cdot(1-\alpha)}$	



[18] Ψώνιος Δ., (1999). «Στατιστική». Εκδόσεις Ζήτη, Θεσσαλονίκη

Παράδειγμα:

Σύμφωνα με τις προδιαγραφές μηχανήματος που παράγει νάϋλον σακούλες, η τυπική απόκλιση της αντοχής τους είναι 1,250 κιλά. Τυχαίο δείγμα από 18 σακούλες έδωσε S=1,9 κιλά. Να ελεγχθεί σε επίπεδο σημαντικότητας α=5% αν το μηχάνημα λειτουργεί με βάση τις προδιαγραφές του.

Απάντηση:

Πριν την διατύπωση των υποθέσεων, σημειώνεται ότι πρόκειται να γίνει έλεγχος για τη διασπορά. Επομένως, $\sigma^2 = 1{,}5625$.

<u>Στάδιο 1ο: Διατύπωση Υποθέσεων</u>

H_0: $\sigma^2 = 1{,}5625$
H_1: $\sigma^2 > 1{,}5625$

<u>Στάδιο 2ο: Στατιστική Συνάρτηση</u>

Η κατανομή της αντοχής για τις νάϋλον σακούλες θεωρείται κανονική. Επιπλέον με δεδομένη την τυχαιότητα του δείγματος και την προϋπόθεση ισχύος της βασικής υπόθεσης, χρησιμοποιούμε την τυχαία μεταβλητή

$X^2_{\nu^*} = \frac{(n-1)\cdot s^2}{\sigma_0^2}$, αφού n=18 < 30

<u>Στάδιο 3ο: Κριτήριο Ελέγχου</u>

Στο δεδομένο επίπεδο σημαντικότητας α=5%, θ' απορρίψουμε την βασική υπόθεση εάν:

$$\chi^2_{\nu^*} > \chi^2_{\nu\cdot\alpha}$$

Στάδιο 4ο: Υπολογισμοί

$$\chi^2_{\nu \cdot \alpha} = \chi^2_{17 \cdot 0,05} = 27{,}59 \text{ (από πίνακες της κατανομής } X^2\text{)}$$

$$\chi^2_{17^*} = \frac{(n-1) \cdot s^2}{\sigma_0^2} = \frac{17 \cdot 3{,}61}{1{,}5625} = 39{,}28$$

Στάδιο 5ο: Στατιστικό Συμπέρασμα

Επειδή $\chi^2_{17^*}$ $(= 39{,}28) > \chi^2_{17 \cdot 0,05}$ $(= 27{,}59)$
απορρίπτουμε τη μηδενική υπόθεση H_0 και δεχόμαστε την H_1 , σε επίπεδο σημαντικότητας α=5%.

2.2.3 ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ ΓΙΑ ΤΗΝ ΑΝΑΛΟΓΙΑ

Στην συνέχεια ακολουθεί ο έλεγχος υποθέσεων για την αναλογία ενός πληθυσμού. Έστω λοιπόν, ότι σε ένα πληθυσμό υπάρχει μία αναλογία ή ένα ποσοστό με μία συγκεκριμένη ιδιότητα. Η αναλογία ορίζεται με p.

Κατόπιν, λαμβάνεται ένα τυχαίο δείγμα από τον πληθυσμό μεγέθους n, του οποίου ένα μέρος x, έχει την εν λόγω ιδιότητα. Η εκτιμήτρια μεταβλητή της p δίνεται από τον λόγο: $\hat{P} = \frac{x}{n}$

Η διενέργεια των ελέγχων υποθέσεων για την αναλογία γίνεται με σκοπό να διαπιστωθεί κατά πόσο η αναλογία αυτή ενός μέρους του πληθυσμού, ισούται με μία ορισμένη αναλογία η οποία συμβολίζεται με P_0.

Όπως συνέβη και σε προηγούμενους ελέγχους υπόθεσης άλλων μεταβλητών, οι πιθανοί συνδυασμοί της μηδενικής υπόθεσης H_0: $p=p_0$ με την εναλλακτική υπόθεση H_1 είναι οι ακόλουθοι:

i. $H_0 : p_0 = p_0$ $\qquad H_1 : p \neq p_0$
ii. $H_0 : p_0 = p_0$ $\qquad H_1 : p > p_0$

iii. H_0 : $p_0 = p_0$ H_1 : $p < p_0$

Προτού γίνει ο έλεγχος των παραπάνω υποθέσεων, σημειώνεται ότι η κατανομή που ακολουθεί η εκτιμήτρια $\widehat{P} = \frac{x}{n}$, είναι κανονική ή κατά προσέγγιση κανονική. Κάτι τέτοιο ισχύει για τον λόγο ότι, τα μεγέθη των δειγμάτων είναι μεγάλα (n > 30) η μέση τιμή της εκτιμήτριας της αναλογίας ισούται με $\mu_{\widehat{P}}$=P και η τυπική απόκλιση υπολογίζεται από την σχέση $S_{\widehat{P}} = \sqrt{\frac{P\,(1-P)}{n}}$. Σημειώνεται ότι κατ' ανάγκη η μεταβλητή P αντικαθίσταται από την τιμή της εκτιμήτριας $\widehat{P}$.

Επομένως, η μεταβλητή

$$Z = \frac{\hat{P}-p}{S_{\widehat{P}}} = \frac{\hat{P}-p_0}{S_{\widehat{P}}}$$

ακολουθεί την τυπική κανονική μεταβλητή.

Για τους ελέγχους των υποθέσεων (i), (ii) και (iii) που καθορίστηκαν παραπάνω, καθώς και για επίπεδο σημαντικότητας α, οι αντίστοιχες περιοχές απόρριψης της μηδενική υπόθεσης, καθορίζονται κατ' αντιστοιχία από τις παρακάτω σχέσεις, όπου $\hat{p}$ είναι η τιμή της εκτιμήτριας $\widehat{P}$.

$$\frac{\hat{p}-p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} > z_{\alpha/2} \quad \text{και} \quad \frac{\hat{p}-p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} < -z_{\alpha/2} \qquad (2.7)$$

$$\frac{\hat{p}-p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} > z_{\alpha} \qquad (2.8)$$

και

$$\frac{\hat{p}-p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} < -z_{\alpha} \qquad (2.9)$$

Ομοίως και σε αυτή τη περίπτωση καθορίζουμε τη κρίσιμη περιοχή, η οποία εξαρτάται από το επίπεδο σημαντικότητας α και στην συνέχεια γίνεται ο υπολογισμός της z, ώστε να διαπιστωθεί κατά πόσο βρίσκεται στην κρίσιμη περιοχή ή όχι. Αν η τιμή της z, εντάσσεται στην κρίσιμη περιοχή, τότε απορρίπτουμε την μηδενική υπόθεση H_0 και αποδεχόμαστε την εναλλακτική υπόθεση H_1. Σε διαφορετική περίπτωση, αποδεχόμαστε την μηδενική υπόθεση H_0.[19]

Παράδειγμα:

Σε ένα δείγμα από 50 κομμάτια περιέχονται 4 ελαττωματικά. Μήπως αυτό αντίκειται στον ισχυρισμό ότι ο πληθυσμός από τον οποίο προέρχεται το δείγμα περιέχει 5% ελαττωματικά, σε επίπεδο σημαντικότητας 5%;

Απάντηση:

Από το παραπάνω πρόβλημα διαπιστώνεται ότι, η μηδενική υπόθεση είναι ότι ο πληθυσμός περιέχει μέχρι 5% ελαττωματικά. Συνεπώς, η υπόθεση η οποία θα πρέπει να ελεγχθεί είναι η εξής:

H_0: p=0,05
H_1: p>0,05

Η εκτίμηση της $\widehat{P}$ που προκύπτει από το συγκεκριμένο δείγμα υπολογίζεται ως εξής: $\widehat{p}$=4/50= 0.08 επίσης, λόγω του ότι, ο αριθμός του δείγματος είναι πολύ μεγάλος n > 30, η μεταβλητή z ακολουθεί κανονική κατανομή.

Επομένως, ο υπολογισμός της μεταβλητής z γίνεται ως εξής:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\hat{p}\left(1-\hat{p}/_n\right)}} => Z = \frac{0.08 - 0.05}{\sqrt{\frac{0.08 \cdot 0.92}{50}}} = \frac{0.03}{0.037} = 0.81$$

---

[19] Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β’ Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη
http://users.auth.gr/dkugiu/Teach/CivilEngineer/hypothesis.pdf

Για την απόρριψη της μηδενική υπόθεσης θα πρέπει η τιμή της μεταβλητής z να είναι μεγαλύτερη του Z_α. από τον πίνακα της κανονικής κατανομής και σε επίπεδο σημαντικότητας α=5%, προκύπτει ότι, $Z < Z_\alpha$ (0,81 < 1,65).

Επομένως, διαπιστώνεται ότι, η παρατηρηθείσα διαφορά δεν είναι στατιστικά σημαντική και δεν συντρέχει λόγος ώστε να απορριφθεί η μηδενική υπόθεση. Με άλλα λόγια, σε επίπεδο σημαντικότητας 5%, δεν υπάρχουν αρκετές ενδείξεις ότι ο πληθυσμός μπορεί να μην περιέχει ελαττωματικά κομμάτια μέχρι 5%.

2.3 ΕΛΕΓΧΟΣ ΠΑΡΑΜΕΤΡΩΝ ΓΙΑ ΔΥΟ ΠΛΗΘΥΣΜΟΥΣ

2.3.1 ΕΛΕΓΧΟΣ ΓΙΑ ΤΗ ΔΙΑΦΟΡΑ ΜΕΣΩΝ ΤΙΜΩΝ

Στην μέχρι τώρα ανάλυση που πραγματοποιήθηκε στο παρόν κεφάλαιο, ο έλεγχος των υποθέσεων αφορούσε την πιθανή διαφορά μεταξύ μίας εκτιμημένης παραμέτρου ενός πληθυσμού και της τιμής της παραμέτρου που θεωρούσαμε γνωστή.

Στη συνέχεια του κεφαλαίου, ο έλεγχος των υποθέσεων θα πραγματοποιηθεί για την περίπτωση που υπάρχουν δύο ομοειδής εκτιμήσεις για μία παράμετρο και επομένως θα πρέπει να εκτιμηθεί η διαφορά τους. Και σε αυτή τη περίπτωση σκοπός είναι να διαπιστωθεί εάν υπάρχει κάποια αιτία που να προκαλεί την συγκεκριμένη διαφορά. Με άλλα λόγια, διερευνάται κατά πόσο η διαφορά των εκτιμήσεων των δύο παραμέτρων οφείλεται στις διακυμάνσεις της δειγματοληψίας ή σε πραγματική διαφορά των αντίστοιχων παραμέτρων των πληθυσμών από τους οποίους λήφθηκαν τα δείγματα.

Στην πραγματικότητα πραγματοποιείται σύγκριση των δύο πληθυσμών χρησιμοποιώντας ως κριτήριο τις μέσες τιμές τους.[20]

Η πρώτη διάκριση γίνεται ανάλογα με το αν οι τυπικές αποκλίσεις του είναι γνωστές ή όχι. Κατόπιν, στην περίπτωση των άγνωστων τυπικών αποκλίσεων, η διάκριση γίνεται ανάλογα με το αν είναι ίσες ή όχι, ή ακόμη αν τα δείγματα είναι μικρά ή μεγάλα.

[20] http://users.auth.gr/dkugiu/Teach/CivilEngineer/hypothesis.pdf

Γνωστές τυπικές αποκλίσεις

Στην παρούσα φάση θα γίνει η υπόθεση ότι, από δύο πληθυσμούς λαμβάνονται δύο μεγάλα δείγματα μεγέθους $n_1>30$ και $n_2>30$ και επιδιώκεται η διακρίβωση εάν οι μέσες τιμές τους διαφέρουν στατιστικά σημαντικά σε ένα προκαθορισμένο επίπεδο σημαντικότητας.

Σημειώνεται ότι, οι δύο πληθυσμοί θα έχουν μέση τιμή μ_1 και τυπική απόκλιση σ_1 και μ_2 και σ_2 αντίστοιχα.

Στον έλεγχο υποθέσεων για την διαφορά των μέσων τιμών, αποσκοπείτε να εξακριβωθεί κατά πόσο οι δύο πληθυσμοί έχουν ίσες μέσες τιμές. Σε περίπτωση που οι μέσες τιμές είναι ίσες μπορεί να εξαχθεί και το συμπέρασμα ότι πρόκειται για ένα πληθυσμό.

Η μηδενική υπόθεση, H_0: $\mu_1=\mu_2$ μπορεί να συνδυαστεί με την εναλλακτική υπόθεση H_1 κατά τους εξής τρόπους:

$H_0 : \mu_1 - \mu_2 = 0$	$H_1: \mu_1 - \mu_2 \neq 0$
$H_0 : \mu_1 - \mu_2 = 0$	$H_1 : \mu_1 - \mu_2 > 0$
$H_0 : \mu_1 - \mu_2 = 0$	$H_1 : \mu_{1-}\mu_2 < 0$

Από τις υποθέσεις που έγιναν προηγουμένως συμπεραίνεται ότι τόσο η μεταβλητή $\overline{\mathrm{X}}_1$, όσο και η μεταβλητή $\overline{\mathrm{X}}_2$, ακολουθούν κανονική κατανομή. Δηλαδή, $\overline{\mathrm{X}}_1 \sim$ N($\bar{\mathrm{x}}_1;\ \mu_1\ , \sigma_1^2/\mathrm{n}$) και $\overline{\mathrm{X}}_2 \sim$ N($\bar{\mathrm{x}}_1;\ \mu_1\ , \sigma_1^2/\mathrm{n}$.

Όσον αφορά την μέση τιμή της μεταβλητής $\overline{\mathrm{X}}_1 - \overline{\mathrm{X}}_2$, αυτή ισούται με τη διαφορά των μέσων τιμών της μεταβλητής $\overline{\mathrm{X}}_1$ και της μεταβλητής $\overline{\mathrm{X}}_2$ και ακολουθεί επίσης την κανονική κατανομή. Επιπλέον, η διασπορά της εν λόγω μεταβλητής, ισούται με το άθροισμα των διασπορών τους, αφού τα δείγματα είναι ανεξάρτητα.

Συνεπώς, η μεταβλητή Ζ, που ορίζεται από την παρακάτω σχέση, ακολουθεί την κατανομή Ν(0,1).

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \qquad (2.10)$$

Αν πάρουμε υπόψη μας ότι υποθέσαμε ότι $\mu_1 = \mu_2$, η σχέση (2.10) μπορεί να διατυπωθεί και ως εξής:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad (2.11)$$

Με βάση τη μεταβλητή Z μπορεί να διενεργηθεί έλεγχος υπόθεσης, με πανομοιότυπο τρόπο που περιγράφηκε σε προηγούμενη παράγραφο.

Ουσιαστικά, η περιοχή απόρριψης της μηδενικής υπόθεσης H_o, σε επίπεδο σημαντικότητας α καθορίζεται από τις σχέσεις (2.1), (2.2) και (2.3), όπου αντί $\frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}}$ θέτουμε την Z, που υπολογίστηκε από την σχέση (2.11).[21]

Άγνωστες ίσες τυπικές αποκλίσεις

Σε περίπτωση που δεν είναι γνωστές οι μεταβλητότητες σ_1^2 και σ_2^2 , τότε λαμβάνονται οι εκτιμήσεις S_1^2 και S_2^2, οπότε η μεταβλητή

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0,1) \qquad (2.13)$$

ακολουθεί επίσης την κατανομή N(0,1). Ο έλεγχος της υπόθεσης εφαρμόζεται όπως και σε προηγούμενους ελέγχους κανονικής κατανομής. Σημειώνεται ότι, τα δείγματα είναι μεγάλα.

[21] Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β’ Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη

Στο σημείο αυτό σημειώνεται ότι, σε περίπτωση που οι πληθυσμοί από τους οποίους προέρχονται τα δείγματα είναι κανονικοί (ή προσεγγιστικά κανονικοί), τα δείγματα είναι μικρά και οι τυπικές αποκλίσεις άγνωστες αλλά ίσες μεταξύ τους, τότε ο έλεγχος της υπόθεσης στην συγκεκριμένη περίπτωση μπορεί να διενεργηθεί μέσω της παρακάτω στατιστική συνάρτησης.

$$t_{n_1+n_2-2} = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, \qquad (2.14)$$

Δηλαδή, μέσω της

$$t_{n_1+n_2-2} = \frac{\overline{X_1}-\overline{X_2}}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \qquad (2.15)$$

όπου t είναι η μεταβλητή της κατανομής του Student με n_1+n_2-2 βαθμούς ελευθερίας και s_p^2 η σταθμισμένη εκτιμήτρια της διασποράς, που βρίσκεται από την παρακάτω σχέση:

$$s_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} \qquad (2.16)$$

Για τον υπολογισμό της σταθμισμένης εκτιμήτριας της διασποράς, λήφθηκε υπόψη ότι, οι μεταβλητότητες των δύο πληθυσμών που έχουν μέσες τιμές μ_1 και μ_2, είναι οι σ_1^2 και σ_2^2.

Για τον λόγο ότι, οι διασπορές σ_1^2 και σ_2^2 είναι ίσες, δίνεται η δυνατότητα χρησιμοποίησης των τιμών του ενός δείγματος. Για παράδειγμα, η εκτίμηση της σταθμισμένης εκτιμήτριας μπορεί να γίνει χρησιμοποιώντας τις τιμές του δείγματος που έχουν μέση τιμή την $\overline{X}_1$ και να εκτιμήσουμε την κοινή μεταβλητότητα ως S_1^2. Κάτι αντίστοιχο μπορεί να γίνει και με το δεύτερο δείγμα και να υπολογιστεί η κοινή μεταβλητότητα ως S_2^2.

Ωστόσο, με αυτό τον τρόπο δεν χρησιμοποιούμε όλες τις πληροφορίες που έχουμε και από τα δύο δείγματα. Για τον λόγο αυτό, χρησιμοποιείται η μεταβλητότητα S^2, η οποία προκύπτει από την σχέση (2.13) ως εκτιμήτρια της σ^2.

Η περιοχή απόρριψης σε επίπεδο σημαντικότητας α καθορίζεται από τις σχέσεις (2.5), (2.6) και (2.7), όπου αντί $\frac{\overline{X}-\mu_0}{s/\sqrt{n}}$ θέτουμε τη μεταβλητή t, που την υπολογίζουμε από τη σχέση (2.15). [22]

Άγνωστες άνισες τυπικές αποκλίσεις

Στην παρούσα περίπτωση οι τυπικές αποκλίσεις των πληθυσμών είναι άγνωστες και τα δείγματα είναι μεγάλα και ανεξάρτητα.

Για την διενέργεια του έλεγχου της υπόθεσης υπολογίζονται οι εκτιμήτριες S_1^{*2} και S_2^{*2} των διακυμάνσεων ${\sigma_1}^2$ και ${\sigma_2}^2$ των δύο πληθυσμών αντίστοιχα.

Η στατιστική συνάρτηση του ελέγχου που χρησιμοποιείται, λόγω του ότι τα δείγματα είναι μεγάλα και βάση του Κεντρικού Οριακού Θεωρήματος, ακολουθεί κατά προσέγγιση την κανονική κατανομή N(0,1).

Η στατική ελέγχου είναι η εξής:

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^{*2}}{n_1} + \frac{S_2^{*2}}{n_2}}} \sim N(0,1) \qquad (2.17)$$

Η διαδικασία του έλεγχου της υπόθεσης διενεργείται όπως και ανωτέρω.

Παράδειγμα:

Για να εκτιμήσουμε την αποτελεσματικότητα ενός προγράμματος εκπαίδευσης εργαζομένων στην εκτέλεση μιας εργασίας εκλέξαμε δυο ομάδες εργαζομένων, από τις οποίες η μία μόνο έκανε τη σχετική εκπαίδευση. Στη συνέχεια εργάστηκαν και οι δυο ομάδες στη συγκεκριμένη εργασία και μετρήθηκαν οι αντίστοιχοι χρόνοι που πέτυχαν. Στην ομάδα που έκανε την εκπαίδευση ήταν 12 εργάτες και στην άλλη 15. Οι μέσες τιμές των χρόνων, που πέτυχαν οι δύο ομάδες ήταν 6,8 min και 9,3 min αντίστοιχα. Αν οι εκτιμήσεις των μεταβλητοτήτων των χρόνων που πέτυχαν οι δύο ομάδες είναι $\mathrm{s}_1^2 =$

[22] http://users.auth.gr/dkugiu/Teach/CivilEngineer/hypothesis.pdf

$10{,}3$ min και $s_2^2 = 15{,}7$ min αντίστοιχα, και υποθέτοντας ότι οι μεταβλητότητες των χρόνων εκτέλεσης μιας εργασίας από εκπαιδευμένους ή όχι εργάτες είναι ίσες, να ελεγχθεί σε επίπεδο σημαντικότητας 5% κατά πόσο υπάρχει διαφορά ικανότητας στις δύο ομάδες των εργαζομένων.

Απάντηση:

Στην περίπτωση αυτή η υπόθεση που θέλουμε να ελέγξουμε είναι η εξής:

$H_0 : \mu_1 - \mu_2 = 0$ και
$H_1 : \mu_1 - \mu_2 \neq 0$

Επειδή $\bar{x}_1 = 6{,}8$, $\bar{x}_2 = 9{,}3$, $s_1^2 = 10{,}3$, $s_2^2 = 15{,}7$, $n_1 = 12$ και $n_2 = 15$ θα υπολογιστεί η σταθμισμένη εκτιμήτρια της διασποράς και στην συνέχεια, λόγω του ότι τα δείγματα είναι μικρά θα υπολογιστεί η μεταβλητή t.

Επομένως, από την σχέση 2.16, έχουμε:

$$s_p = \sqrt{\frac{11 \cdot 10{,}3 + 14 \cdot 15{,}7}{25}} = 3{,}65.$$

Οι βαθμοί ελευθερίας είναι 12+15-2=25. Για 25 βαθμούς ελευθερίας και επίπεδο σημαντικότητας α=5% βρίσκουμε ότι $t_{25,0.025}=2{,}059$.

Επίσης, από τη σχέση 2.15 η μεταβλητή t υπολογίζεται ως εξής:

$$t = \frac{6{,}8 - 9{,}3}{3{,}65\sqrt{\frac{1}{12} + \frac{1}{15}}} = -1{,}77.$$

Επειδή $-2{,}059 < -1{,}77 < 2{,}059$ δεχόμαστε την υπόθεση H_0 σε επίπεδο σημαντικότητας 5%. Δηλαδή, δεν υπάρχει ένδειξη ότι η εκπαίδευση επιδρά στην ικανότητα των εργαζομένων να εκτελέσουν τη συγκεκριμένη εργασία.

Παράδειγμα:

Αναλυτής αγοράς ερευνά τη συμπεριφορά δύο τύπων εταιρειών όσον αφορά τη στρατηγική διαφήμισής τους. Μία από τις ερευνώμενες μεταβλητές είναι το ποσό των διαφημιστικών τους δαπανών την προηγούμενη χρονιά. Επέλεξε γι' αυτό δύο τυχαία και ανεξάρτητα δείγματα από κάθε τύπο εταιρείας, τα οποία έδωσαν τα παρακάτω αποτελέσματα.

Τύπος Α ετ.	$n_A = 60$ $\overline{X}_A = 10.500$ € $S_A^2 = 1.290$ €
Τύπος Β ετ.	$n_B = 70$ $\overline{X}_B = 10.200$ € $S_B^2 = 956$ €

Μπορούμε να συμπεράνουμε απ' αυτά τα δείγματα ότι ο τύπος Α των εταιρειών δαπανά, κατά μέσο όρο μεγαλύτερα ποσά για διαφήμιση, απ' ότι ο τύπος Β; το επίπεδο σημαντικότητας είναι α=5%.

Απάντηση :

Στάδιο 1^o: Διατύπωση Υποθέσεων :

H_0: $\mu_A - \mu_B = 0$
H_1: $\mu_A - \mu_B > 0$ μονόπλευρος πάνω

Στάδιο 2^o: Στατιστική Συνάρτηση:

Από το συγκεκριμένο παράδειγμα είναι λογικό να υποθέσουμε ότι και για τους δύο τύπους εταιρειών η δαπάνη για διαφήμιση συνιστά τυχαία μεταβλητή η οποία κατανέμεται κανονικά. Επιπλέον αφού $v = n_1 + n_2 - 2 = 128 > 30$ χρησιμοποιούμε για τον έλεγχο την μεταβλητή Ζ που ακολουθεί την κανονική κατανομή:

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

<u>Στάδιο 3^o: Κριτήριο Ελέγχου:</u>

Αφού μας δίδεται επίπεδο σημαντικότητας α=5%, θα απορρίψουμε τη μηδενική υπόθεση εάν $Z > Z_\alpha$

<u>Στάδιο 4^o: Υπολογισμοί:</u>

$$Z_a = Z_{0,05} = 1{,}645$$

$$Z = \frac{10{,}5 - 10{,}2}{\sqrt{\frac{1{,}29}{60} + \frac{0{,}956}{70}}} = 1{,}60$$

<u>Στάδιο 5^o: Στατιστικό Συμπέρασμα</u>

Αφού Z (=1,60) < $Z_{0,05}$(=1,645) αποδεχόμαστε την μηδενική υπόθεση H_0 σε επίπεδο σημαντικότητας α = 5%.

Επομένως, συμπεραίνεται ότι, ο τύπος των εταιρειών Α δεν δαπανά, κατά μέσο όρο μεγαλύτερα ποσά για διαφήμιση, απ' ότι ο τύπος εταιρειών Β.

2.3.2 ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ ΓΙΑ ΤΟ ΛΟΓΟ ΔΙΑΣΠΟΡΩΝ

Η διατύπωση των υποθέσεων για τον έλεγχο των διασπορών των δύο κανονικών πληθυσμών γίνεται, κατά τα γνωστά, ανάλογα τη μορφή του έλεγχου, ως εξής:

$H_0: \frac{\sigma_1^2}{\sigma_2^2} = 1$ $\qquad$ $H_0: \sigma_1^2 = \sigma_2^2$

ή

$H_1: \frac{\sigma_1^2}{\sigma_2^2} \neq 1$ $\mathrm{H}_1: \sigma_1^2 \neq \sigma_2^2$

Σύμφωνα με την κατανομή δειγματοληψίας του λόγου δύο διασπορών, εάν από δύο πληθυσμούς των X και Y [$X \sim N(\mu_X, \sigma_X^2)$ και $Y \sim N(\mu_Y, \sigma_Y^2)$] πάρουμε δύο τυχαία και ανεξάρτητα δείγματα μεγέθους n_X και n_Y και υπολογίσουμε τις διακυμάνσεις τους S_X^2 και S_Y^2, αποδεικνύεται ότι η αναλογία των δύο σχετικών διασπορών, που είναι μια τιμή της τυχαίας μεταβλητής F^*, ακολουθεί την F με ν_1 (= $n_X - 1$) και ν_2 (= $n_Y - 1$) βαθμούς ελευθερίας, αντίστοιχα.

Τα παραπάνω παρουσιάζονται και μέσα από την σχέση 2.18 που ακολουθεί:

$$F^*_{(\nu_1,\nu_2)} \equiv \frac{\frac{S_X^2}{\sigma_X^2}}{\frac{S_Y^2}{\sigma_Y^2}} \sim F_{(\nu_1,\nu_2)} \qquad (2.18)$$

όπου: ν_1, ν_2 οι βαθμοί ελευθερίας του αριθμητή και παρανομαστή αντίστοιχα.

Σε πολλές περιπτώσεις, οι διασπορές των δύο πληθυσμών σ_X^2 και σ_Y^2 αντίστοιχα, μπορεί να είναι άγνωστες. Για τον λόγο αυτό, πραγματοποιείται συνήθως η υπόθεση ότι, οι διασπορές των δύο πληθυσμών είναι ίσες μεταξύ τους και ισούται με σ^2.

Δηλαδή, $\sigma_X^2 = \sigma_Y^2 = \sigma^2$. Επομένως, η παραπάνω σχέση μπορεί να γραφεί ως εξής:[23]

$$F^*_{(\nu_1,\nu_2)} \equiv \frac{S_X^2}{S_Y^2} \sim F_{(\nu_1,\nu_2)} \qquad (2.19)$$

Παράδειγμα:

Πολλά μέλη Δ.Ε.Π. κάποιου Πανεπιστημίου ισχυρίζονται ότι πρέπει να καταργηθούν οι εμβόλιμες περίοδοι εξετάσεων. Ένα από τα επιχειρήματα τους είναι ότι

[23] Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β' Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη

δεν έχουν παρατηρήσει σημαντική διαφορά στις διακυμάνσεις της βαθμολογίας των φοιτητών μεταξύ κανονικής και εμβόλιμης, εξεταστικών περιόδων. Κάποιος καθηγητής μάλιστα μελέτησε δύο ανεξάρτητα και τυχαία δείγματα των επιδόσεων των φοιτητών σε κάθε μία περίοδο, με τα παρακάτω αποτελέσματα.

Να ελεγχθεί σε α = 2% εάν η παρατηρηθείσα δειγματική διαφορά των διακυμάνσεων είναι στατιστικά σημαντική.

Κανονική Περίοδος Εξετάσεων	$n_1 = 25$	$S_1^2 = 4{,}84$
Εμβόλιμη Περ. Εξετάσεων	$n_2 = 16$	$S_2^2 = 9{,}61$

Απάντηση:

Στάδιο 1ο : Διατύπωση Υποθέσεων:

$$H_0: \frac{\sigma_2^2}{\sigma_1^2} = 1$$

$$H_1: \frac{\sigma_2^2}{\sigma_1^2} \neq 1$$

Στάδιο 2ο : Στατιστική Συνάρτηση:

Θεωρούμε ότι οι τυχαίες μεταβλητές της βαθμολογίας των φοιτητών σε κάθε περίοδο ακολουθούν την κανονική και με την προϋπόθεση ότι ισχύει η βασική υπόθεση, η στατιστική ελέγχου ακολουθεί την $F_{(\nu_2,\nu_1)}$ κατανομή.

Στάδιο 3ο : Κριτήριο Ελέγχου:

Επειδή $s_2^2 > s_1^2$ στο δεδομένο α = 2% θα απορρίψουμε την H_0 αν $F^*_{(\nu_2,\nu_1)} = \frac{s_2^2}{s_1^2} >$ $F_{(\nu_2,\nu_1).\,\alpha/2}$

$\nu_2 = n_2 - 1 = 15$

$\nu_1 = n_1 - 1 = 24$

Επομένως η H_0 απορρίπτεται αν :

$$F^*_{(15,24)} > F_{(15,24).0,01}$$

Στάδιο 4ο : Υπολογισμοί :

$F_{(15,24).0,01} = 2,89$

$F^* = \dfrac{9,61}{4,84} = 1,99$

Στάδιο 5ο : Στατιστικό Συμπέρασμα:

Επειδή $F^*(= 1,99) < F_{(15,24)}$(=2,89) δεν μπορεί ν' απορριφθεί η H_0, σε α =2%.

Με άλλα λόγια επιβεβαιώνεται στατιστικά η άποψη ότι, σε επίπεδο σημαντικότητας 2%, δεν χρειάζονται οι εμβόλιμες εξεταστικές περίοδοι.

2.3.3 ΕΛΕΓΧΟΣ ΓΙΑ ΤΗΝ ΔΙΑΦΟΡΑ ΑΝΑΛΟΓΙΩΝ

Έστω λοιπόν, ότι $\widehat{P}_1$ και $\widehat{P}_2$ είναι οι εκτιμήτριες των αναλογιών δύο δειγμάτων, τα οποία έχουν μέγεθος n_1 και n_2 αντίστοιχα και αναλογίες p_1 και p_2.

Για την πραγματοποίηση του ελέγχου, θα θεωρηθεί ως μηδενική υπόθεση ότι, δεν υφίσταται διαφορά ανάμεσα στις αναλογίες των πληθυσμών, δηλαδή ότι $p_1 = p_2$. Με την συγκεκριμένη υπόθεση, θεωρείται ότι τα δείγματα ουσιαστικά προέρχονται από τον ίδιο πληθυσμό.

Για να ελέγξουμε την υπόθεση χρειαζόμαστε την κατανομή της διαφοράς των δύο αναλογίων των δύο πληθυσμών. Για τον λόγο αυτό θα πρέπει να υπολογιστούν οι ακόλουθες σχέσεις:

$$\mu_{p_1-p_2} = \mu_{p_1} - \mu_{p_2} = p_1 - p_2 \qquad (2.20)$$

και

$$S_{p_1-p_2} = \sqrt{S^2_{p_1} + S^2_{p_2}} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}. \qquad (2.21)$$

Εφόσον η μηδενική υπόθεση είναι η H_0: $p_1 = p_2$, δηλαδή ότι πρόκειται για τον ίδιο πληθυσμό, μπορούμε να αντικαταστήσουμε στις σχέσεις (2.20) και (2.21) ως εξής:

$$p_1 = p_2 = p,$$

$$\mu_{p_1-p_2} = 0,$$

και

$$S_{p_1-p_2} = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad (2.22)$$

Ως $\boldsymbol{\hat{p}}$ χρησιμοποιούμε μια εκτίμηση της αναλογίας του πληθυσμού, που τη βρίσκουμε από την σχέση:

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2} \qquad (2.23)$$

Σε περίπτωση που τα μεγέθη των δειγμάτων είναι αρκετά μεγάλα, η μεταβλητή:

$$Z = \frac{\hat{P}_1-\hat{P}_2-0}{S_{p_1-p_2}} = \frac{\hat{P}_1-\hat{P}_2}{S_{p_1-p_2}} \sim N(0,1) \qquad (2.24)$$

ακολουθεί την κατανομή N(0,1).

Με βάση τα παραπάνω μπορούμε εύκολα να ελέγξουμε τη μηδενική υπόθεση, όπως κάναμε και σε προηγούμενους ελέγχους.[24]

Παράδειγμα:

Μια υφαντουργική βιομηχανία, προκειμένου να προγραμματίσει μαζική παραγωγή ενός είδους υφάσματος, έστειλε δείγματα από το ίδιο ύφασμα, αλλά με διαφορετικά χρώματα, σε διαφόρους εμπόρους. Συγκεκριμένα, έστειλε 200 δείγματα κόκκινου και 200 πράσινου χρώματος σε εμπόρους της Αθήνας και της Θεσσαλονίκης. Αποτέλεσμα αυτού ήταν να πάρει 130 παραγγελίες από εκείνους που πήραν κόκκινα δείγματα και 100 παραγγελίες από άλλους που πήραν πράσινα δείγματα. Να ελεγχθεί κατά πόσον υπάρχει ένδειξη ότι τα συγκεκριμένα χρώματα επηρεάζουν τις πωλήσεις.

Απάντηση:

Η υπόθεση την οποία πρέπει να ελέγξουμε εδώ είναι η εξής:

$$H_0: p_\kappa = p_\pi$$

$$H_1: p_\kappa \neq p_\pi.$$

Δηλαδή τα χρώματα δεν επηρεάζουν τις πωλήσεις του συγκεκριμένου προϊόντος

Εφόσον υποθέτουμε ότι $p_\kappa = p_\pi$, το ίδιο ισχύει και για τις αναλογίες $\hat{p}_\kappa$ και $\hat{p}_\pi$.

Η εκτίμηση του **p,** από τα $\hat{p}_\kappa$ και $\hat{p}_\pi$, δίνεται από τη σχέση:

$$\hat{p} = \frac{n_\kappa \cdot \hat{p}_\kappa + n_\pi \cdot \hat{p}_\pi}{n_\kappa + n_\pi} = \frac{200(0{,}5 + 0{,}65)}{400} = \frac{1{,}15}{2} = 0{,}575$$

Συνεπώς,

$$Z = \frac{(\hat{p}_\kappa - \hat{p}_\pi)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_\kappa} + \frac{1}{n_\pi}\right)}} = \frac{0{,}65 - 0{,}50}{\sqrt{0{,}57 \cdot 0{,}43 \cdot \frac{2}{200}}} = 3{,}02$$

[24] Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β’ Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη

Εφόσον η τιμή του $\mathbf{Z}$ είναι μεγαλύτερη του $Z_{-\alpha/2} = Z_{0,975} = 1{,}96$, μπορούμε να συμπεράνουμε ότι υπάρχει ένδειξη ότι υφίσταται διαφορά στις πωλήσεις εξαιτίας του χρώματος και επομένως απορρίπτουμε τη μηδενική υπόθεση για επίπεδο σημαντικότητας 5%.

2.4 ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο παρόν κεφάλαιο παρουσιάστηκε ο τρόπος με τον οποίο διενεργείται ο έλεγχος των υποθέσεων για διάφορες παραμέτρους του πληθυσμού, όπως είναι η μέση τιμή, η διασπορά και η αναλογία. Στην συγκεκριμένη περίπτωση πρόκειται για έλεγχο υποθέσεων που προκύπτει από την διαφορά που υπάρχει μεταξύ της εκτίμησης για μία από αυτές τις παραμέτρους και της τιμής που θεωρούταν γνωστή για αυτήν.

Στην συνέχεια, έλεγχος υποθέσεων διενεργήθηκε και για την περίπτωση που υπήρχαν ομοειδής εκτιμήσεις για μία παράμετρο. Στους συγκεκριμένους ελέγχους, στόχος είναι η εξακρίβωση της διαφοράς μεταξύ των εκτιμήσεων. Οι έλεγχοι υποθέσεων που αναλύθηκαν στην συγκεκριμένη περίπτωση αφορούσαν την διαφορά των μέσων, την διαφορά των διασπορών και την διαφορά των αναλογιών μεταξύ δύο πληθυσμών.

ΚΕΦΑΛΑΙΟ ΤΡΙΤΟ
ΕΦΑΡΜΟΓΕΣ ΤΩΝ ΕΛΕΓΧΩΝ ΥΠΟΘΕΣΕΩΝ

3.1 ΠΕΔΙΟ ΕΦΑΡΜΟΓΗΣ ΕΛΕΓΧΩΝ ΥΠΟΘΕΣΕΩΝ

Πέρα όμως από τους παραπάνω κλάδους, η στατιστική συμπερασματολογία βρίσκει πολύ εφαρμογή και στην οικονομική επιστήμη. Χαρακτηριστικό είναι το γεγονός ότι, δείκτες όπως είναι το γενικό επίπεδο τιμών, της απασχόλησης, της παραγωγικότητας, της συνολικής δαπάνης, κ.λπ., αποτελούν συχνά αντικείμενο στατιστικής επεξεργασίας και διενέργειας στατιστικών υποθέσεων.[25]

3.2 Ο ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ ΣΤΗΝ ΟΙΚΟΝΟΜΙΚΗ ΕΠΙΣΤΗΜΗ

Στην οικονομική επιστήμη, οι θεωρίες που αναπτύσσονται εκφράζουν σχέσεις μεταξύ δύο ή περισσότερων μεταβλητών και προσδιορίζουν την συμπεριφορά ενός φαινομένου, μέσα από την επεξεργασία ορισμένων υποθέσεων.

Χαρακτηριστικό παράδειγμα αποτελεί η θεωρία της ζήτησης των αγαθών. Στην συγκεκριμένη περίπτωση, οι οικονομολόγοι έχουν δημιουργήσει ορισμένες υποθέσεις για να καταλήξουν στο συμπέρασμα ότι, η ζητούμενη ποσότητα για κάθε αγαθό, εξαρτάται (αποτελεί συνάρτηση) από την τιμή του αγαθού, τις τιμές των άλλων αγαθών, καθώς και το εισόδημα των καταναλωτών. Οι υποθέσεις για την δημιουργία της θεωρίας της ζήτησης είναι ότι, οι καταναλωτές έχουν σαφείς προτιμήσεις βασισμένες στο εισόδημα που διαθέτουν, ενώ επίσης, επιδιώκουν την μεγιστοποίηση της χρησιμότητας τους από την απόκτηση κάθε αγαθού.[26]

Πέρα όμως από την διατύπωση της θεωρίας της ζήτησης των αγαθών, θα πρέπει να ελεγχθεί ο βαθμός στον οποίο αυτή ανταποκρίνεται στην πραγματικότητα. Τόσο ο εμπειρικός έλεγχος των θεωριών, όσο και η άσκηση οικονομικής πολιτικής, η διενέργεια

---

[25] Begg David, Fischer Stanley, Dornbusch Rudiger (2006). «Εισαγωγή στην οικονομική». Εκδόσεις Κριτική, Αθήνα

[26] Τζαβαλής Ηλίας, (2008). «Οικονομετρία». Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών

προβλέψεων, κλπ., προϋποθέτουν την ποσοτικοποίηση των οικονομικών σχέσεων που δημιουργούνται.

Σε αυτό συμβάλλει η οικονομετρία, η οποία αποτελεί έναν συνδυασμό μεταξύ της οικονομικής επιστήμης, των μαθηματικών και της στατιστικής. Ένας από τους βασικούς στόχους της οικονομετρίας είναι ο έλεγχος της σημαντικότητας των οικονομικών θεωριών.

Ο έλεγχος των υποθέσεων στην οικονομετρία εφαρμόζεται μέσα από την παλινδρόμηση. Για τους οικονομολόγους, η παλινδρόμηση αποτελεί το σημαντικότερο κεφάλαιο της στατιστικής επιστήμης. Ο λόγος είναι ότι στα οικονομικά, η παλινδρόμηση επιλέγεται και εφαρμόζεται σχεδόν πάντοτε σε σχέση με τις υπόλοιπες ποσοτικές μεθόδους. Σημειώνεται ότι στόχος της παλινδρόμησης είναι η εκτίμηση των σχέσεων μεταξύ των οικονομικών μεταβλητών, όπως είναι για παράδειγμα η συνάρτηση της ζήτησης ενός αγαθού, όπως αναφέρθηκε και προηγουμένως.[27]

Στο σημείο αυτό αξίζει να σημειωθεί ότι ακολουθούν δύο παραδείγματα εφαρμογής της στατιστικής συμπερασματολογίας στην παλινδρόμηση. Το ένα αναφέρεται σε ένα απλό διμεταβλητό υπόδειγμα, ενώ το δεύτερο αναφέρεται σε ένα πολλαπλό υπόδειγμα παλινδρόμησης.

Απλό υπόδειγμα παλινδρόμησης

Παράδειγμα

Έστω ότι η καθαρή επένδυση (I) και η μεταβολή στις πωλήσεις (S), συνδέονται με την σχέση $I = \beta_0 + \beta_1 S + u$. Ένα τυχαίο δείγμα από πέντε επιχειρήσεις έδωσε τα αποτελέσματα του ακόλουθου πίνακα.

Επιχειρήσεις	I σε χιλ. ευρώ	S σε χιλ. ευρώ
A	10	30
B	24	20

[27] Βενέτης Ιωάννης, (2009). «Εισαγωγικές διαλέξεις στην Οικονομετρία». Εκδόσεις Γκιούρδας, Αθήνα
Κιντής Α., (2010). «Σύγχρονη οικονομετρική ανάλυση - Τόμος Α'». Εκδόσεις Gutenberg, Αθήνα

Γ	41	70
Δ	14	50
Ε	6	-20

Να διερευνηθεί εάν η σχέση μεταξύ των επενδύσεων και της μεταβολής των πωλήσεων είναι στατιστικά σημαντική, σε επίπεδο σημαντικότητας α=5%.

Επιπλέον, δίνεται τα εξής:

$\Sigma xy=1380$

$\Sigma x^2=4600$

$\Sigma x^2=784$

$\bar{Y} = 30$

$\bar{X} = 19$

Για λόγους ευκολίας των υπολογισμών, όπου Ι θέτουμε Υ και όπου S θέτουμε Χ.

Απάντηση

Αρχικά από τα παραπάνω θα πρέπει να υπολογιστούν οι συντελεστές β_o και β_1. Με την μέθοδο των ελαχίστων τετραγώνων προκύπτει ότι, οι εκτιμημένοι συντελεστές υπολογίζονται ως εξής:

$$\hat{\beta}_1 = \frac{\Sigma xy}{\Sigma x^2} = \frac{1380}{4600} = 0{,}3$$

και

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 19 - 0{,}3x30 = 10$$

Στην συνέχεια, διατυπώνονται η μηδενική και η εναλλακτική υπόθεση.

H_0: $\beta_1=0$
H_1: $\beta_1>0$

Η στατιστική που θα χρησιμοποιηθεί για τον έλεγχο της υπόθεσης είναι η στατιστική F. Ο υπολογισμός της στατιστικής F γίνεται από τον ακόλουθο τύπο:

$$\mathrm{F} = \frac{\hat{\Sigma}_y^2 / 1}{\hat{\Sigma}_u^2 / n-2} \sim \mathrm{F}_{1,n-2}$$

Ο αριθμητής υπολογίζεται ως εξής:

$$\hat{\Sigma}_y^2 = \hat{\beta}_1 \Sigma xy = 0{,}3\ x\ 1380 = 414$$

Ο υπολογισμός του παρανομαστή γίνεται ως εξής:

$$\hat{\sigma}^2 = \frac{\hat{\Sigma}_u^2}{n-2} = \frac{\Sigma_y^2 - \Sigma_{\hat{y}}^2}{n-2} = \frac{784-414}{5-2} = 123{,}333$$

Κάνοντας αντικατάσταση στην σχέση της F, υπολογίζεται ότι,

$$F = \frac{414}{123,333} = 3,357$$

Επειδή, $F = 3,357 < F_{1,3,0,05} = 10,1$ δεν απορρίπτουμε την μηδενική υπόθεση H_0. Αυτό σημαίνει ότι, η μεταβλητή πωλήσεις δεν ερμηνεύει σημαντικό ποσοστό της μεταβλητότητας της καθαρής επένδυσης και επομένως δεν μπορεί να απορριφθεί ότι, ο συντελεστής β_1 είναι μηδέν.

Πολλαπλό υπόδειγμα παλινδρόμησης

Παράδειγμα

Έστω ότι σε ένα δείγμα από 33 παρατηρήσεις έχουν γίνει οι εξής υπολογισμοί για τις μεταβλητές Y, X_1 και X_2.

$\bar{Y} = 19,5$

$\overline{X_1} = 6$

$\overline{X_2} = 10$

$\Sigma y_2 = 3$

$\Sigma \chi_1^2 = 30$

$\Sigma \chi_2^2 = 80$

$\Sigma \chi_1 y = 7$

$\Sigma \chi_2 y = 4$

$\Sigma \chi_1 \chi_2 = 40$

Επιπλέον, έστω ότι η εκτίμηση της παλινδρόμησης

$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + u$ είναι η εξής:

$\bar{Y} = 18,5 + 0,5X_1 - 0,2X_2$

Να διενεργηθεί έλεγχος της υπόθεσης ότι β_1=0 σε επίπεδο σημαντικότητας α=5%.

Δίνεται επίσης,

$$(X'X)^{-1} = \begin{pmatrix} 0{,}1 & -0{,}05 \\ -0{,}05 & 0{,}0375 \end{pmatrix}$$

Απάντηση

Για να διενεργηθεί ο έλεγχος της υπόθεσης θα πρέπει να υπολογιστεί μέσω της στατιστικής t-student ο ακόλουθος λόγος

$$t = \frac{\widehat{\beta_1} - \beta_1}{\widehat{Se}_{\widehat{\beta}1}} \sim t_{n-2,a/2}$$

Από τα δεδομένα του παραδείγματος ο συντελεστής β_1, καθώς και ο εκτιμημένος συντελεστής β_1 είναι γνωστοί. Χρειάζεται λοιπόν να υπολογιστεί ο παρανομαστής της στατιστικής t, ο οποίος αποτελεί το τυπικό σφάλμα.

Ο υπολογισμός του τυπικού σφάλματος γίνεται ως εξής:

$$\widehat{Se}_{\widehat{\beta}1} = \sqrt{\widehat{Var}_{\widehat{\beta 1}}} = \sqrt{\sigma^2 (X'X)^{-1}}$$

όπου

$$\sigma^2 = \frac{\Sigma \hat{u}^2}{n-k} = \frac{\Sigma_y^2 - \hat{\beta}_1 \Sigma \chi_1 y - \hat{\beta}_2 \Sigma \chi_2 y}{n-k}$$

Με αντικατάσταση των δεδομένων από την εκφώνηση, προκύπτει ότι

$$\sigma^2 = \frac{3 - 0{,}5 \times 7 - (-0{,}2) \times 0{,}4}{33 - 3} = \frac{0{,}3}{30} = 0{,}01$$

Στην συνέχεια υπολογίζεται η σχέση $\sigma^2(X'X)^{-1}$. Από τα δεδομένα η μήτρα $(X'X)^{-1}$ είναι γνωστή, επομένως ο υπολογισμός έχει ως εξής:

$$\sigma^2(X'X)^{-1} = 0{,}1\begin{pmatrix} 0{,}1 & -0{,}05 \\ -0{,}05 & 0{,}0375 \end{pmatrix}$$

Άρα,

$$\widehat{Var}_{\widehat{\beta\varepsilon\tau}} = \begin{pmatrix} 0{,}001 & -0{,}0005 \\ -0{,}0005 & 0{,}000375 \end{pmatrix}$$

όπου, $\widehat{Var}_{\widehat{\beta\varepsilon\tau}}$ είναι η μήτρα διακυμάνσεων – συνδιακυμάνσεων του $\widehat{\beta}_{\varepsilon\tau}$

Επίσης, από την θεωρία της οικονομετρίας γνωρίζουμε ότι, [28]

$$\widehat{Var}_{\widehat{\beta\varepsilon\tau}} = \begin{pmatrix} \widehat{Var}_{\widehat{\beta 1}} & \widehat{Cov}_{(\widehat{\beta 1},\widehat{\beta 2})} \\ \widehat{Cov}_{(\widehat{\beta 1},\widehat{\beta 2})} & \widehat{Var}_{\widehat{\beta 2}} \end{pmatrix}$$

Επομένως,

$$\widehat{Se}_{\widehat{\beta}1} = \sqrt{\widehat{Var}_{\widehat{\beta 1}}} = \sqrt{0{,}001} = 0{,}0316$$

Στο σημείο αυτό, ο έλεγχος υπόθεσης γίνεται ως εξής:

$$t = \frac{\widehat{\beta_1} - \beta_1}{\widehat{Se}_{\widehat{\beta}1}} = \frac{0{,}5 - 0}{0{,}0316} = 15{,}8 > t_{31,0,025} \approx 1{,}96$$

Εφ' όσον 15,8>1,96 απορρίπτουμε την μηδενική υπόθεση ότι, ο συντελεστής β_1 ισούται με το μηδέν. Συμπεραίνεται λοιπόν, ότι όταν η X_2 παραμένει σταθερή και μεταβάλλεται η X_1, τότε η μεταβολή αυτή θα έχει σημαντική επίδραση στην Υ.[29]

---

[28] Αγιακλόγλου Χρήστος, Μπένος Θεοφάνης, (2007). «Εισαγωγή στην οικονομετρική ανάλυση». Εκδόσεις Μπένου

[29] Τσιώνας Βασίλειος, (2009). «Στατιστική με εφαρμογές στην Οικονομετρία». Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών

Βάμβουκας Γεώργιος, (2007). «Σύγχρονη Οικονομετρία». Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών

3.3 ΕΥΡΥΤΕΡΟ ΠΕΔΙΟ ΕΦΑΡΜΟΓΗΣ ΤΩΝ ΕΛΕΓΧΩΝ ΥΠΟΘΕΣΕΩΝ

Το κομμάτι της στατιστικής συμπερασματολογίας, πέρα από τον κλάδο της οικονομικής επιστήμης, χρησιμοποιείται σε διάφορους ακόμη τομείς, μερικοί εκ των οποίων είναι οι εξής:

- Βιοστατιστική
- Στατιστικός Έλεγχος Ποιότητας
- Κοινωνική Στατιστική
- Στατιστικές μέθοδοι στον Επιχειρηματικό Σχεδιασμό
- Στατιστικές μέθοδοι στα Χρηματοοικονομικά [30]

Βιοστατιστική

Η βιοστατιστική σαν έννοια προσδιορίζει την εφαρμογή των στατιστικών μεθόδων στις επιστήμες της βιολογίας και της ιατρικής. Η επιστήμη της βιολογίας και της ιατρικής, αποτελούν επιστήμες οι οποίες παρέχουν πλήθος ερευνητικών υποθέσεων και σεναρίων. Για τον λόγο αυτό, καθίσταται αναγκαία η εφαρμογή σε αυτές της στατιστικής συμπερασματολογίας.

Η βιοστατιστική, σαν εργαλείο της στατιστικής και συγκεκριμένα του στατιστικού ελέγχου υποθέσεων, συμπεριλαμβάνει κάθε ποσοτική προσέγγιση που μπορεί να χρησιμοποιηθεί για να δώσει απαντήσεις σε ερευνητικά ερωτήματα ή να προβεί σε έλεγχο της ορθότητας διάφορων επιστημονικών θεωριών, υποθέσεων και σεναρίων. Στο σημείο αυτό σημειώνεται ότι, η συνηθέστερη εφαρμογή των στατιστικών μεθόδων που εφαρμόζεται στις επιστήμες της υγείας, είναι ο σχεδιασμός και η ανάλυση των κλινικών δοκιμών.

Η εκπαίδευση και εξειδίκευση πάνω στον τομέα της Βιοστατιστικής παρέχεται κυρίως μέσω μεταπτυχιακών προγραμμάτων. Παρόλα αυτά, εκπαιδευτικά σεμινάρια για την βιοστατιστική και την χρησιμότητά της παρέχονται και σε σχολές δημόσιας υγείας,

[30] Κολύβα-Μαχαίρα Φ., Μπόρα-Σέντα Ε., (1998). «Στατιστική θεωρία και εφαρμογές». Εκδόσεις Ζήτη, Θεσσαλονίκη

σχολές που συνδέονται με την δασοπονία ή την γεωργία, καθώς επίσης και στα τμήματα Στατιστική επιστήμης, όπως είναι προφανές. [31]

Όπως ήδη αναφέρθηκε, εκπαιδευτικά προγράμματα της βιοστατιστικής συνδέονται με τις οικολογικές επιστήμες. Στόχος των μαθημάτων βιοστατιστικής που διενεργούνται στις συγκεκριμένες σχολές είναι η εισαγωγή εννοιών και μεθόδων για τον έλεγχο υποθέσεων για μία ή και περισσότερες μεταβλητές ενός δείγματος, ή ομάδων δειγμάτων. Στην συγκεκριμένη περίπτωση, πολλές φορές τα εκπαιδευτικά μαθήματα της βιοστατιστικής συνδυάζονται με μαθήματα Πειραματικού Σχεδιασμού.

Στο σημείο αυτό σημειώνεται ότι, η πλειοψηφία των επιστημόνων της στατιστικής θεωρούν ότι η διάκριση της στατιστικής και της βιοστατιστικής δεν μπορεί να υφίσταται για τον λόγο ότι, ένας στατιστικός έχει την δυνατότητα να επιλύσει προβλήματα ιατρικής φύσεως που άπτονται των στατιστικών ελέγχων υποθέσεων.

Ωστόσο, στην πραγματικότητα κάτι τέτοιο δεν μπορεί να πραγματοποιηθεί. Μέσα από έρευνες που έχουν πραγματοποιηθεί τόσο σε χώρες της Ευρώπης, όσο και τις Ηνωμένες Πολιτείες Αμερικής, έχει διαπιστωθεί ότι παρόλο που είναι ειδικευμένοι της στατιστικής επιστήμης, δεν μπορούν να διενεργήσουν ελέγχους υποθέσεων στην ιατρική επιστήμη και την επιστήμη της βιολογίας, καθώς απαιτούνται εξειδικευμένες γνώσεις πάνω στην ιατρική επιστήμη που αυτοί δεν γνωρίζουν. Οι γνώσεις αυτές αναφέρονται τόσο σε θέματα ορολογίας όσο και σε θέματα ανάλυσης.

Ένα απλό παράδειγμα χρήσης των ελέγχων υποθέσεων μέσα στον τομέα της ιατρικής επιστήμης είναι το ακόλουθο και σχετίζεται με την διερεύνηση της σχέσης μεταξύ της χρήσης κινητών τηλεφώνων και της συχνότητας εμφάνισης όγκου στον εγκέφαλο.

Ο έλεγχος της υπόθεσης στην προκειμένη περίπτωση περιλαμβάνει τις εξής υποθέσεις: η μηδενική υπόθεση θεωρεί ότι δεν υπάρχει σχέση μεταξύ της χρήσης κινητών τηλεφώνων και της συχνότητας εμφάνισης όγκου στον εγκέφαλο. Η εναλλακτική υπόθεση στο εν λόγω πρόβλημα είναι ότι υπάρχει σχέση μεταξύ της χρήσης κινητών τηλεφώνων και της συχνότητας εμφάνισης όγκου στον εγκέφαλο. Παρόλα αυτά,

[31] Τριχόπουλος Δ., Τζώνου Α., Κατσουγιάννη κ., (2001). «Βιοστατιστική». Εκδόσεις Παρισιάνου, Αθήνα

δεν προσδιορίζεται κατά πόσο η σχέση αυτή είναι θετική ή αρνητική. Στην συγκεκριμένη περίπτωση, ο έλεγχος υπόθεσης που θα διενεργηθεί είναι δίπλευρος.

Σε περίπτωση που ο έλεγχος υπόθεσης που θα διενεργηθεί είναι μονόπλευρος έλεγχος προς τα πάνω, τότε η μηδενική υπόθεση θα διατηρηθεί ίδια, δηλαδή, η χρήση κινητών τηλεφώνων δεν επηρεάζει τη συχνότητα εμφάνισης όγκου στον εγκέφαλο, ενώ η εναλλακτική υπόθεση θα δημιουργηθεί ως εξής: η χρήση των κινητών τηλεφώνων αυξάνει τη συχνότητα εμφάνισης όγκου στον εγκέφαλο. Μέσα από τις εν λόγω υποθέσεις, θεωρείται ότι η συχνότητα εμφάνισης όγκου στον εγκέφαλο είναι μεγαλύτερη σε εκείνους που χρησιμοποιούν κινητά.

Στην συνέχεια και με βάση τα δεδομένα της μελέτης πραγματοποιείται ο στατιστικός έλεγχος της υπόθεσης και κατόπιν αποδεχόμαστε ή απορρίπτουμε την μηδενική υπόθεση. Ο τρόπος διενέργειας και επιλογής απόρριψης ή όχι της μηδενικής υπόθεσης έχει αναλυθεί σε προηγούμενο κεφάλαιο της παρούσας πτυχιακής εργασίας.

Στο σημείο αυτό αξίζει να σημειωθεί ότι, σε περίπτωση που η μηδενική υπόθεση απορριφθεί, τότε εξάγεται το συμπέρασμα ότι σύμφωνα με τα δεδομένα της παρούσας μελέτης υπάρχει σχέση μεταξύ της χρήσης κινητών και της συχνότητας εμφάνισης όγκου στον εγκέφαλο. Σε διαφορετική περίπτωση, εάν δηλαδή αποδεχθούμε την μηδενική υπόθεση, τότε δεν εξάγεται το συμπέρασμα ότι δεν υπάρχει σχέση μεταξύ των δύο μεταβλητών, χρήση κινητού τηλεφώνου και συχνότητας εμφάνισης όγκου στον εγκέφαλο, αλλά συμπεραίνεται ότι, τα δεδομένα της εν λόγω μελέτης δεν πρόσφεραν σημαντικές ενδείξεις για την απόρριψη της μηδενικής υπόθεσης. Επομένως, είναι πολύ πιθανό η απόρριψη της μηδενικής υπόθεσης να επέλθει με τα δεδομένα που θα παρέχει μία άλλη μελέτη. [32]

Στατιστικός Έλεγχος Ποιότητας

Ουσιαστικά, ο Στατιστικός Έλεγχος Ποιότητας αποτελεί ένα σύνολο στατιστικών μεθόδων το οποίο παρέχει πληροφορίες για μία επιχείρηση σχετικά με τις προδιαγραφές, την παραγωγή και τον έλεγχο των παραγόμενων προϊόντων της.

[32] http://www.mednet.gr/archives/2010-4/pdf/691.pdf

Στο σημείο αυτό σημειώνεται ότι, ως ποιότητα ορίζεται το σύνολο των χαρακτηριστικών που επιθυμεί η επιχείρηση να έχει κάθε προϊόν που παράγει. Τα χαρακτηριστικά αυτά διαφέρουν από πελάτη σε πελάτη και σχετίζονται με την απόδοση του προϊόντος, την διάρκεια ζωής του προϊόντος, τις δυνατότητες του, κ.λπ. [33]

Ο Στατιστικός Έλεγχος Ποιότητας διακρίνεται σε τρία επιμέρους υποσύνολα, τα οποία περιλαμβάνουν στατιστικές μεθόδους, οι οποίες αναφέρονται σε διαφορετικές φάσεις της παραγωγικής διαδικασίας. Τα υποσύνολα αυτά είναι τα εξής:

- Σχεδιασμός και Ανάλυση Πειραμάτων (Design of Experiments): περιέχει στατιστικές τεχνικές και μεθόδους που συμβάλλουν στην σχεδίαση της παραγωγικής διαδικασίας.
- Στατιστικός Έλεγχος Διεργασιών (Statistical Process Control): πρόκειται για την διαδικασία του ελέγχου που πραγματοποιείται κατά την διάρκεια της παραγωγής των προϊόντων.
- Δειγματοληψία Αποδοχής (Acceptance Sampling): μέσω της χρήσης στατιστικών μεθόδων, καθορίζει εάν μία παρτίδα προϊόντων θα απορριφθεί ή θα γίνει αποδεκτή και θα προωθηθεί προς τον πελάτη.

Στο σημείο αυτό σημειώνεται ότι, ένα γνωστό στατιστικό πακέτο μέσω του οποίου γίνεται η στατιστική ανάλυση των δεδομένων είναι το πακέτο “MINITAB”.

Κοινωνική στατιστική

Η κοινωνική στατιστική αποτελεί ένα ακόμη πεδίο εφαρμογής των ελέγχων των υποθέσεων. Στόχος της κοινωνικής στατιστικής είναι η ποσοτική διερεύνηση υποθέσεων, που σχετίζονται με θέματα όπως είναι τα ακόλουθα:

- Θέματα υγείας
- Κοινωνικής πρόνοιας
- Εκπαίδευσης και απασχόλησης
- Μετανάστευσης
- Κοινωνικής και επαγγελματικής κινητικότητας

[33] Ατζουλάκος Δ., (2008). «Στατιστικός Έλεγχος Ποιότητας – Β’ Έκδοση». Σημειώσεις παραδόσεων. Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς

Στο σημείο αυτό αξίζει να σημειωθεί ότι, ως κοινωνική στατιστική, ορίζεται η μέθοδος μέσα από την οποία τα δεδομένα που προέρχονται από το κοινωνικό πλαίσιο να μπορούν να αναλυθούν και να αξιολογηθούν. [34]

3.4 Η ΑΝΑΛΥΣΗ ΤΗΣ ΔΙΑΚΥΜΑΝΣΗΣ (ANALYSIS OF VARIANCE – ANOVA)

Η ανάλυση της διακύμανσης αποτελεί μία ιδιαίτερη στατιστική μέθοδο, κατά την οποία, η μεταβλητότητα ενός συνόλου (πληθυσμού) διασπάται σε μικρότερα δείγματα, ώστε να μπορεί να διαπιστωθεί η σημαντικότητα των πηγών προέλευσής τους.

Η θεωρία της ανάλυσης της διακύμανσης αναπτύχθηκε από τον άγγλο στατιστικό Sir Ronald Aylmer Fisher και αποτελεί ένα πεδίο της σύγχρονης στατιστικής θεωρίας.[35]

Η ανάλυση της διακύμανσης έχει αναπτυχθεί με σκοπό να μπορεί να διενεργηθεί ένας στατιστικός έλεγχος υποθέσεων για τις παραμέτρους που έχουν παρουσιαστεί σε προηγούμενο κεφάλαιο, για περισσότερους από δύο πληθυσμούς.

Προτού, γίνει η παρουσίαση ενός παραδείγματος για την ανάλυση της διακύμανσης, αξίζει να σημειωθεί ότι η ANOVA παρουσιάζει ένα πρόβλημα κατά την εφαρμογή της, το οποίο συνήθως επιλύεται με την χρήση του κατάλληλου λογισμικού πακέτου. Το συγκεκριμένο πρόβλημα αναφέρεται στο πως μπορεί να γίνει η σύγκριση μεταξύ κ ίδιων παραμέτρων, όταν το κ είναι μεγαλύτερο από 2. Σημειώνεται ότι, ως κ θεωρούνται οι ίδιες παράμετροι των πληθυσμών.

Μία λύση αποτελεί η χρήση του παραγοντικού και η σύγκριση των παραμέτρων ανά δύο. Δηλαδή, οι έλεγχοι που πρέπει να διενεργηθούν είναι, $c=\binom{k}{2}=\frac{k\,(k-1)}{2}$. Σημειώνεται μία τέτοια διαδικασία είναι χρονοβόρα. Κάτι τέτοιο συμπεραίνεται από το γεγονός ότι, αν για παράδειγμα το κ είναι 4 (δηλαδή σχετικά μικρό), οι έλεγχοι (c) που πρέπει να πραγματοποιηθούν είναι 6. Στην περίπτωση αυτή, όσο περισσότερες είναι οι παρατηρήσεις κ, τόσο περισσότεροι έλεγχοι θα πρέπει να πραγματοποιηθούν.

---

[34] Κολύβα-Μαχαίρα Φ., Μπόρα-Σέντα Ε., (1998). «Στατιστική θεωρία και εφαρμογές». Εκδόσεις Ζήτη, Θεσσαλονίκη

[35] http://www.aua.gr/gpapadopoulos/files/anova12-13a.pdf

Στην συνέχεια, μέσω του παραδείγματος που ακολουθεί, θα πραγματοποιηθεί ο εξής έλεγχος ανάλυσης διακύμανσης, για επίπεδο σημαντικότητας α%.

Η μηδενική και εναλλακτική υπόθεση ορίζονται ως εξής:

H_0: $\mu_i=\mu_j$

H_1: $\mu_i \neq \mu_j$, για ένα τουλάχιστον ζεύγος (i,j)

Παράδειγμα:

Μία βιομηχανία αυτοκινήτων έχει κατασκευάσει ένα μοντέλο αυτοκινήτου του οποίου τα χαρακτηριστικά είναι η άνεση, η καλύτερη ποιότητα που παρέχει και η χαμηλότερη τιμή του, έναντι των ανταγωνιστριών βιομηχανιών. Η διεύθυνση του τμήματος μάρκετινγκ της εν λόγω εταιρείας, προώθησε ένα πείραμα δε τρεις διαφορετικές περιοχές προκειμένου να αποφασίσει με ποιον τρόπο θα γίνει η προώθηση του συγκεκριμένου μοντέλου.

Προέβη λοιπόν, στο εξής πείραμα: στην πρώτη περιοχή η διαφήμιση του αυτοκινήτου θα έδινε έμφαση στην άνεση, στην δεύτερη περιοχή θα δινόταν έμφαση στην ποιότητα κατασκευής και τέλος, στην τρίτη περιοχή η έμφαση θα δινόταν στην τιμή πώλησης.

Στην συνέχεια καταγράφονται τα αποτελέσματα των πωλήσεων με βάση την κάθε περιοχή για το διάστημα των 5 πρώτων μηνών.

Μήνας	Πωλήσεις			
	Περιοχή 1 (Άνεση)	Περιοχή 2 (Ποιότητα)	Περιοχή 3 (Τιμή)	
1	86	90	82	
2	79	76	68	

3	81	88	73	
4	70	82	71	
5	84	89	81	
Μέσος όρος	80	85	75	80

Η διεύθυνση του τμήματος μάρκετινγκ ενδιαφέρεται να μάθει, σε επίπεδο σημαντικότητας 5%, εάν υπάρχουν διαφορές στις πωλήσεις που προέκυψαν από τις τρεις διαφημιστικές στρατηγικές και τα πειράματα που διενεργήθηκαν.[36]

Απάντηση:

Ο έλεγχος υπόθεσης που θα διενεργηθεί είναι ο εξής:

H_0: $\mu_1=\mu_2=\mu_3$
H_1: $\mu_1 \neq \mu_2 \neq \mu_3$, για ένα τουλάχιστον ζεύγος (i,j)

Από το παραπάνω παράδειγμα διαπιστώνεται ότι κ = 3 ανεξάρτητα τυχαία δείγματα. Επιπλέον, η υπόθεση που πραγματοποιείται είναι ότι και τα τρία δείγματα προέρχονται από κανονικό πληθυσμό με διακύμανση σ^2, η οποία είναι κοινή και στους τρείς πληθυσμούς. Επίσης, n=5

Μέσα από τον έλεγχο υπόθεσης θέλουμε να ελέγξουμε εάν σε επίπεδο σημαντικότητας 5%, οι πωλήσεις των αυτοκινήτων της εταιρείας επηρεάζονται από τις διαφορετικές στρατηγικές που η ίδια έχει ακολουθήσει.

Στο σημείο αυτό και πριν κατασκευαστεί ο πίνακας ανάλυσης διακύμανσης, θα πρέπει να υπολογιστούν τα ακόλουθα αθροίσματα τετραγώνων:

Το άθροισμα (G – Grand total) του συνόλου των παρατηρήσεων (ν=15), δίνεται από τον εξής τύπο:

[36] http://www.aua.gr/gpapadopoulos/files/anova12-13a.pdf

$$G = \sum_{ij} y_{ij} = \sum_{j=1}^{3} T_j = 1200$$

Το ολικό άθροισμα των τετραγώνων SS_{Total} έχει ν-1 βαθμούς ελευθερίας και υπολογίζεται από τον ακόλουθο τύπο:

$$SS_{Total} = \sum_{ij} y_{ij}^2 - \frac{G^2}{n} = 698$$

με ν-1 = 14 βαθμούς ελευθερίας.

Το άθροισμα των τετραγώνων μεταξύ των δειγμάτων υπολογίζεται ως εξής:

$$SSTr = \sum_{j=1}^{3} \frac{T_j^2}{n_j} - \frac{G^2}{\nu} = 250$$

με κ-1 = 2 βαθμούς ελευθερίας.

Τέλος, το άθροισμα των τετραγώνων του σφάλματος (άθροισμα τετραγώνων μέσα στα δείγματα), ορίζεται ως εξής:

$$SS_{Total} = SSTr + SSE$$

$$SSE = SS_{Total} - SSTr = 448$$

με ν-κ = 15-3 = 12 βαθμούς ελευθερίας.

Για την κατασκευή του πίνακα ανάλυσης της διακύμανσης θα πρέπει να υπολογιστούν, το μέσο άθροισμα των τετραγώνων μεταξύ των δειγμάτων (MSTr) και το μέσο άθροισμα των τετραγώνων μέσα στα δείγματα (MSE). Ο υπολογισμός των εν λόγω τύπων γίνεται ως εξής:

$$MSTr = \frac{SSTr}{k-1} = 125$$

και

$$MSE = \frac{SSE}{\nu - \kappa} = 37{,}33$$

Τέλος, υπολογίζεται ο λόγος των διακυμάνσεων, μέσω του οποίου διενεργείται ο έλεγχος της υπόθεσης. Ο λόγος των διακυμάνσεων είναι ο ακόλουθος:

$$F_{Tr} = \frac{MSTr}{MSE} = 3{,}35$$

και ακολουθεί την κατανομή F με κ-1 και ν-κ βαθμούς ελευθερίας. Δηλαδή,

$$F_{Tr} = \frac{MSTr}{MSE} \sim F_{k-1;\nu-\kappa}$$

Στην συνέχεια ακολουθεί ο πίνακας ανάλυσης διακύμανσης.

Πηγή μεταβλητότητας	Βαθμοί ελευθερίας	Άθροισμα τετραγώνων	Μέσο άθροισμα τετραγώνων	Κριτήριο F
Μεταξύ των δειγμάτων	2	250	125	3,35
Σφάλμα μέσα στα δείγματα	12	448	37,33	
Σύνολο	14	698		

Για επίπεδο σημαντικότητας 5%, $F_{Tr} = \frac{MSTr}{MSE} \sim F_{k-1;\nu-\kappa;\alpha} = 3{,}88$. Για την απόρριψη της μηδενικής υπόθεσης θα πρέπει να ισχύει:

$$F_{Tr} > F_{k-1;\nu-\kappa;\alpha}$$

Εφ' όσον 3,35 < 3,88, δεν μπορούμε να απορρίψουμε την μηδενική υπόθεση H_0, με βάση τα δεδομένα του δείγματος και επομένως, συμπεραίνεται ότι, δεν υπάρχουν διαφορές μεταξύ των πωλήσεων αυτοκινήτων και των περιοχών – στρατηγικών που αναμένεται να ακολουθήσει η εταιρεία.

ΚΕΦΑΛΑΙΟ 4: ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ

Αφού ολοκληρλωθηκε η παράθεση της βιβλιογραφίας στα προηγούμενα κεφάλαια, παρουσιάζεται μια μελέτη περίπτωσης, ενός παραδείγματος ελέγχου υποθέσεων. Το παράδειγμα αυτά αφορά στον έλεγχο υποθέσεων, μέσω στατιστικής ανάλυσης, μιας ιδιωτικής επιχείρησης. Πιο συγκεκριμένα το παράδειγμα αυτό αναλύεται παρακάτω.

Μια ιδιωτική εταιρεία απασχολεί 100 υπαλλήλους. Σκοπός της παρούσας μελέτης είναι η συγκέντρωση στοιχείων σχετικά με τα χαρακτηριστικά των υπαλλήλων, μέσω χορήγησης ερωτηματολογίων και η ανάλυση των δεδομένων αυτών μέσω χρήσης του στατιστικού πακέτου MINITAB. Πιο συγκεκριμένα θα γίνει έλεγχος υποθέσεων για έναν πληθυσμό: μέσης τιμής, διασποράς και αναλογίας με τη χρήση κατάλληλων τεστ (t-test). Επίσης θα γίνει έλεγχος υποθέσεων δύο πληθυσμών. Με αυτόν τον τρόπο θα γίνει εφαρμογή της θεωρίας που αναφέρθηκε στα προηγούμενα κεφάλαια. Τα δεδομένα και οι μεταβλητές δίνονται αναλυτικά στο Παράρτημα, στο τέλος της εργασίας.

Ακολουθεί, λοιπόν, η ανάλυση των δεδομένων ξεκινόντας από την περιγραφική ανάλυση των μεταβλητών της παρούσας εργασίας. Εν συνεχεία ακολουθεί ο έλεγχος υποθέσεων κάποιων μεταβλητών.

ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

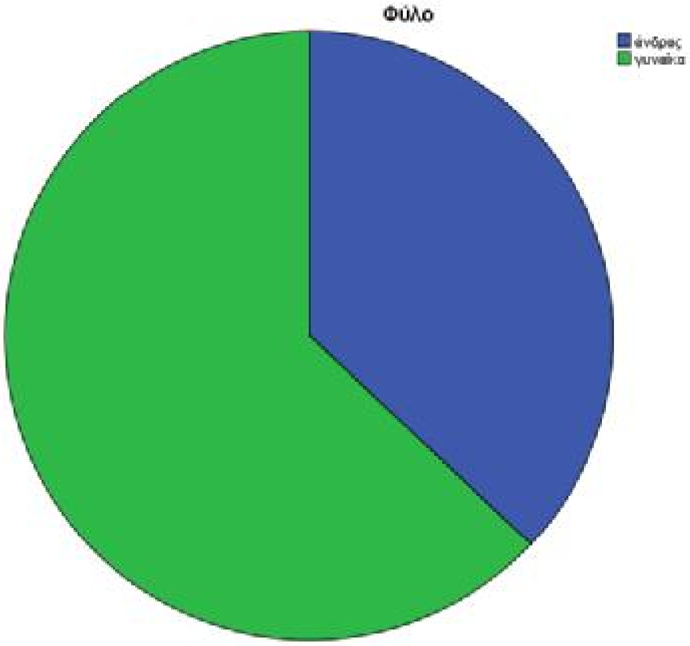
Ι. ΠΕΡΙΓΡΑΦΙΚΗ ΑΝΑΛΥΣΗ

Στη συνέχεια σας παραθέτουμε την περιγραφική ανάλυση της κάθε μιας μεταβλητής. Αρχικα παρουσιάζονται τα αποτελέσματα σε πίνακα και εν συνεχεία σε σχήμα ή θηκόγραμμα. Η ανάλυση των αποτελεσμάτων ξεκινά με την παρουσίαση της πρώτης μεταβλητής του δείγματος (Ν = 100) που είναι το φύλο.

Πρώτη μεταβλητή: ΦΥΛΟ

Φύλο			
	Συχνότητα	**Ποσοστό**	**Συνολικό ποσοστό**
άνδρας	37	37,0	37,0
γυναίκα	63	63,0	100,0
ΣΥΝΟΛΟ	**100**	**100,0**	

Το μεγαλύτερο ποσοστό των εργαζομένων είναι γυναίκες σε ποσοστό 63% και το 37% είναι άντρες.



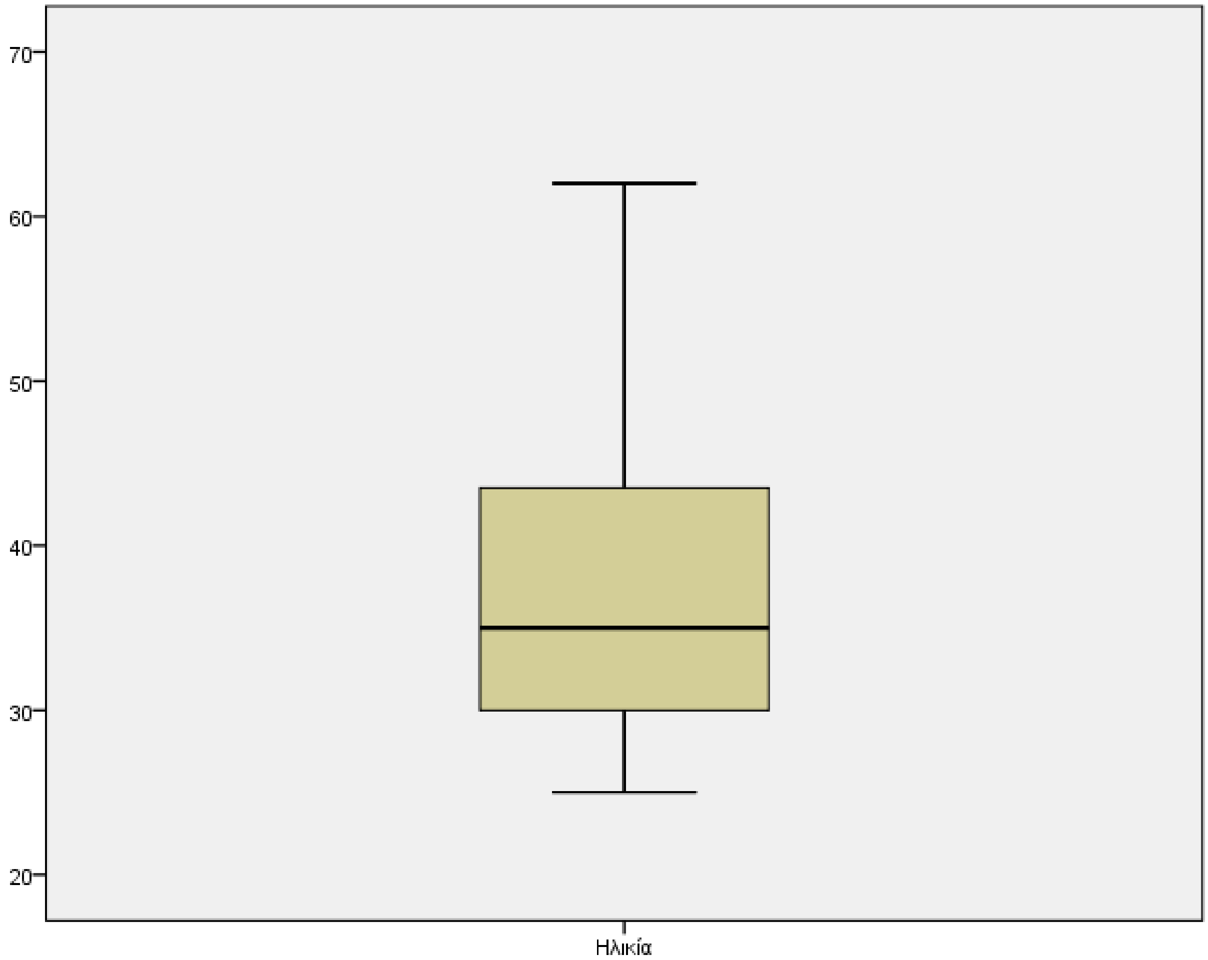
Δεύτερη μεταβλητή: ΗΛΙΚΙΑ

Ηλικία											
	N	Εύρος	Μικρότε-ρη τιμή	Μεγαλύ-τερη τιμή	Μέση τιμή	Τυπική απόκλιση	Διακύ μανση	Λοξότητα		Κυρτότητα	
Ηλικία	100	37	25	62	37,84	10,086	101,7 32	,906	,241	-,035	,478
N (δείγμα)	**100**										

Οι ερωτώμενοι έχουν μέση ηλικία 37,84 ετών, με μικρότερη ηλικία αυτή των 25 ετών και μεγαλύτερη ηλικία αυτή των 62 ετών. Το εύρος της ηλικίας είναι 37 ετών και η τυπική απόκλιση 10,086. Ακολουθεί αναλυτικός πίνακας με τα ποσοστά της κάθε ηλικίας των εργαζομένων.

Ηλικία				
Τιμές		Συχνότητα	Ποσοστό	Συνολικό ποσοστό
	25	5	5,0	5,0
	26	4	4,0	9,0
	27	2	2,0	11,0
	28	3	3,0	14,0
	29	4	4,0	18,0
	30	15	15,0	33,0
	31	1	1,0	34,0
	32	6	6,0	40,0
	33	4	4,0	44,0
	34	1	1,0	45,0
	35	6	6,0	51,0
	36	3	3,0	54,0
	38	8	8,0	62,0
	39	5	5,0	67,0
	40	1	1,0	68,0
	41	2	2,0	70,0
	42	1	1,0	71,0
	43	4	4,0	75,0
	44	3	3,0	78,0
	45	2	2,0	80,0
	46	4	4,0	84,0
	48	2	2,0	86,0
	50	1	1,0	87,0
	53	1	1,0	88,0
	54	1	1,0	89,0
	55	2	2,0	91,0
	58	2	2,0	93,0
	60	5	5,0	98,0
	62	2	2,0	100,0
	ΣΥΝΟΛΟ	**100**	**100,0**	

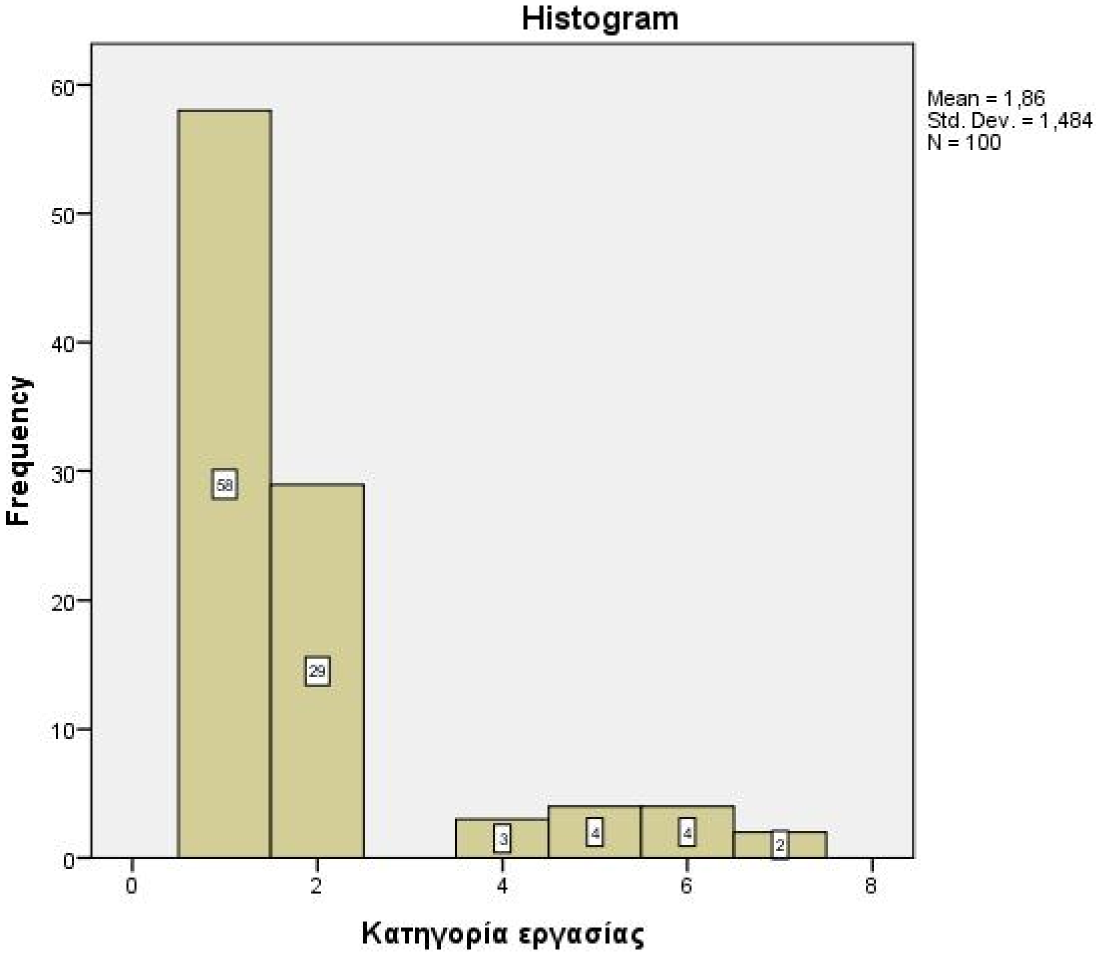
Παρακάτω ακολουθεί το θηκόγραμμα της μεταβλητής ηλικία.



Τρίτη μεταβλητή: ΚΑΤΗΓΟΡΙΑ ΕΡΓΑΣΙΑΣ

Κατηγορία εργασίας			
	Συχνότητα	**Ποσοστό**	**Συνολικό ποσοστό**
υπάλληλος γραφείου	58	58,0	58,0
εκπαιδευόμενος υπάλληλος	29	29,0	87,0
πτυχιούχος	3	3,0	90,0
πτυχιούχος με ΜΔΕ	4	4,0	94,0
χρηματ. αναλυτής	4	4,0	98,0
προϊστάμενος	2	2,0	100,0
ΣΥΝΟΛΟ	**100**	**100,0**	

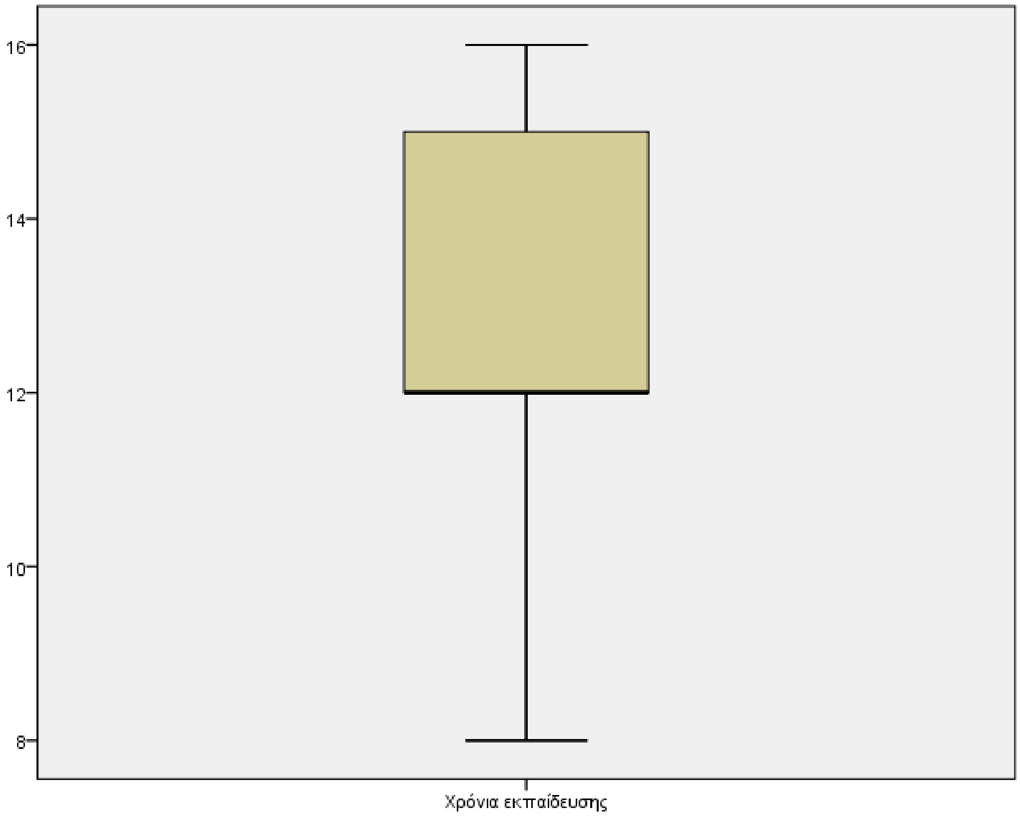
Το μεγαλύτερο ποσοστό των εργαζομένων (58%) είναι υπάλληλοι γραφείου και ένα μεγάλο ποσοστό (29%) είναι εκπαιδευόμενοι υπάλληλοι. Μικρότερα ποσοστά των εργαζομένων είναι πτυχιούχοι με ή χωρίς μεταπτυχιακό τίτλο σπουδών, χρηματικοί αναλυτές. Το μικρότερο ποσοστό, της τάξεως του 2% των εργαζομένων, είναι προϊστάμενοι. Ακολουθεί και ιστόγραμμα με τα αποτελέσματα αυτά.



Τέταρτη μεταβλητή: ΧΡΟΝΙΑ ΕΚΠΑΙΔΕΥΣΗΣ

Χρόνια εκπαίδευσης											
	N	Εύρος	Μικρότε-ρη τιμή	Μεγαλύ-τερη τιμή	Μέση τιμή	Τυπική απόκλιση	Διακύ μανση	Λοξότητα		Κυρτότητα	
Χρόνια εκπαίδ.	100	8	8	16	12,48	2,435	5,929	-,530	,241	-,519	,478
N	**100**										

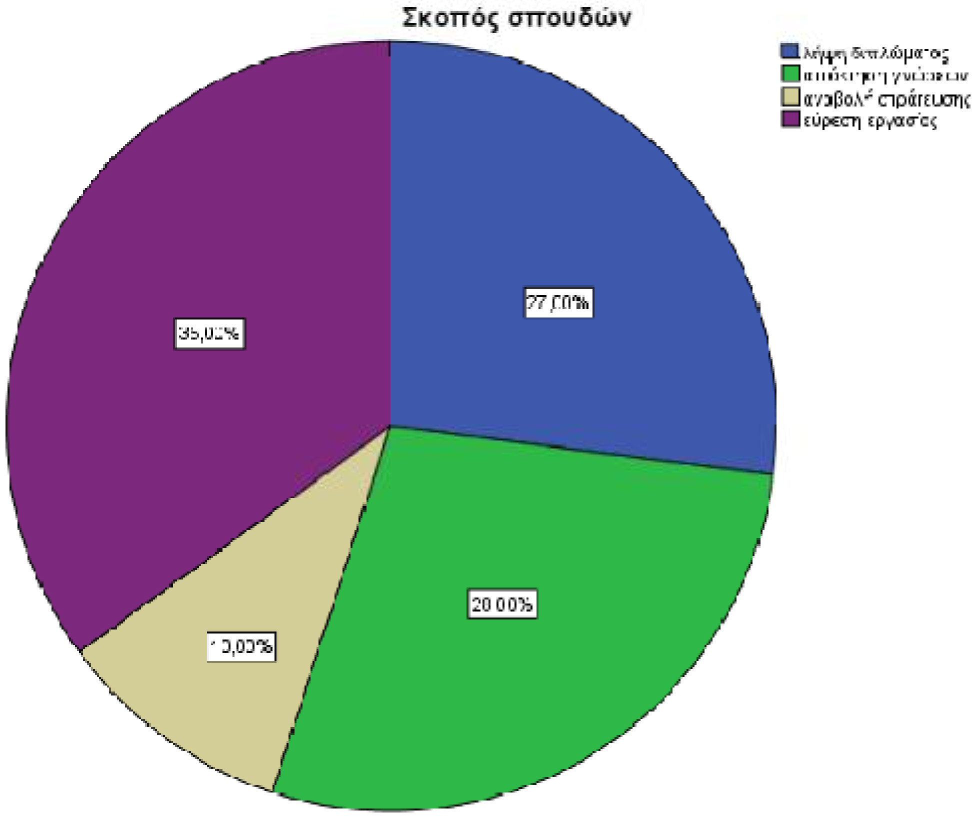
Η μέση τιμή του χρόνου εκπαίδευσης των εργαζομένων της εταιρείας είναι 12, 48 έτη με χαμηλότερα έτη τα 8 και περισσότερα τα 16. Το εύρος είναι 8 έτη και η τυπική απόκλιση 2,435. Ακολουθεί το θηκόγραμμα της συγκεκριμένης μεταβλητής.



Πέμπτη μεταβλητή: ΣΚΟΠΟΣ ΣΠΟΥΔΩΝ

Σκοπός σπουδών			
	Συχνότητα	**Ποσοστό**	**Συνολικό ποσοστό**
λήψη διπλώματος	27	27,0	27,0
απόκτηση γνώσεων	28	28,0	55,0
αναβολή στράτευσης	10	10,0	65,0
εύρεση εργασίας	35	35,0	100,0
ΣΥΝΟΛΟ	**100**	**100,0**	

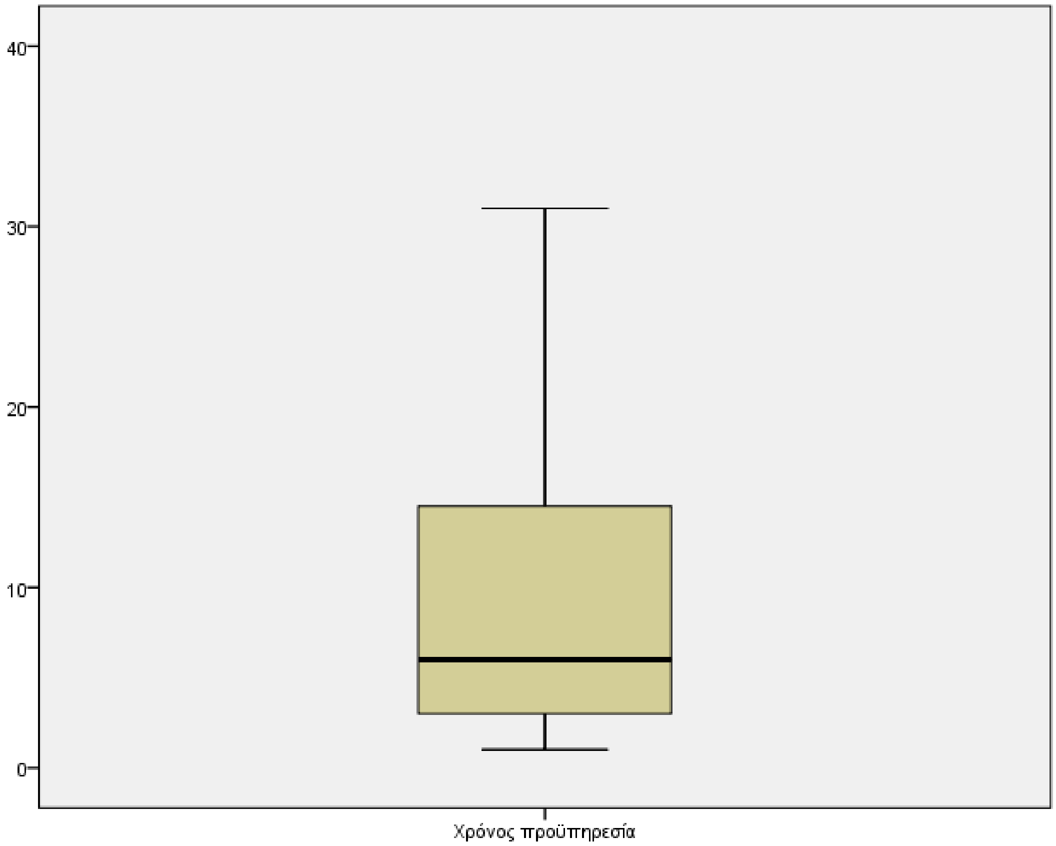
Ο σκοπός σπουδών των περισσότερων εργαζομένων είναι η εύρεση εργασίας σε ποσοστό 35%. Ένα μεγάλο ποσοστό των εργαζομένων (28%) σπούδασε για την απόκτηση γνώσεων και για την λήψη ενός διπλώματος (27%). Τέλος, ένα μικρό ποσοστό του δείγματος, σπούδασε με σκοπό την αναβολή στράτευσης (10%). Ακολουθεί και διάγραμμα με τα παραπάνω ποσοστά.



Έκτη μεταβλητή: ΈΤΗ ΠΡΟΫΠΗΡΕΣΙΑΣ

Έτη προϋπηρεσίας										
	N	Εύρος	Μικρότερη τιμή	Μεγαλύτερη τιμή	Μέση τιμή	Τυπική απόκλιση	Διακύμανση	Λοξότητα		Κυρτότητα	
Έτη προϋπηρεσίας	100	30	1	31	9,48	8,601	73,969	1,033	,241	-,123	,478
N	**100**										

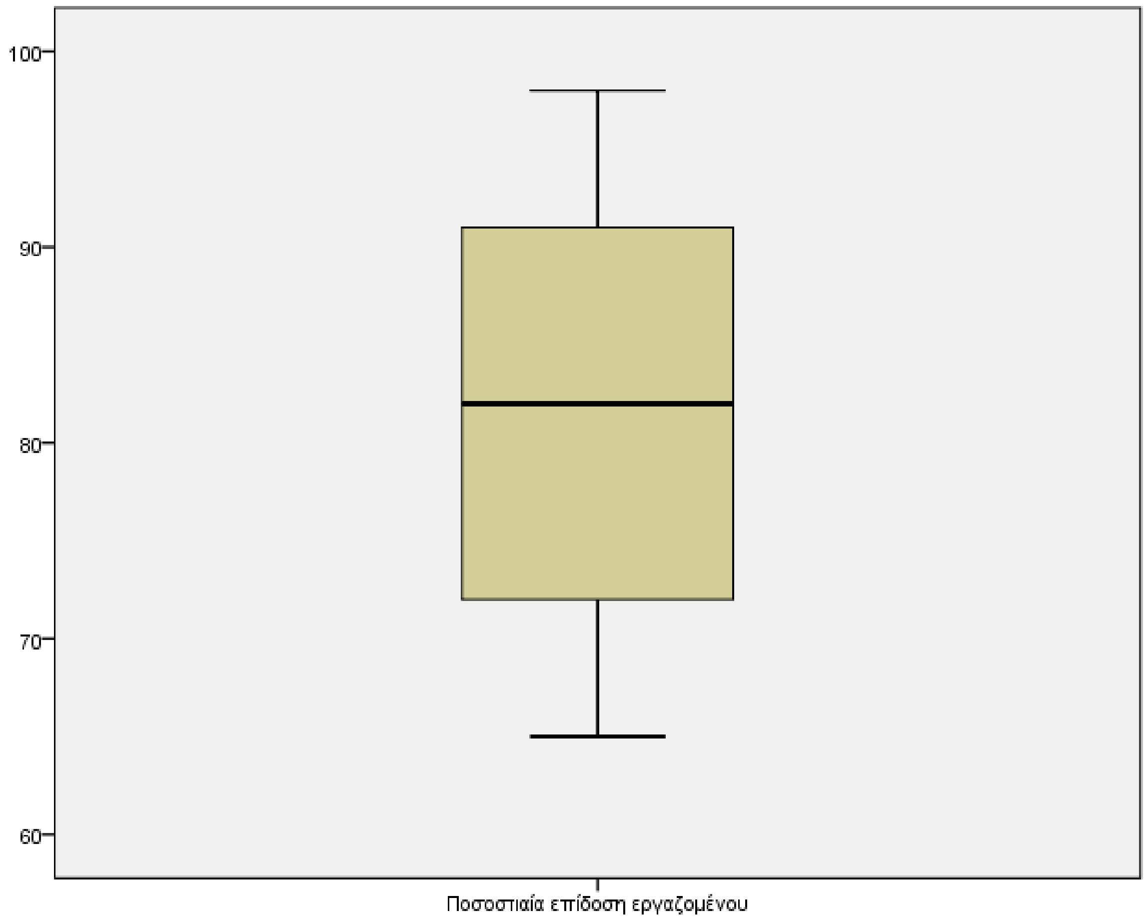
Τα μέσα έτη προϋπηρεσίας των εργαζομένων είναι 9,48. Η χαμηλότερη προϋπηρεσία των εργαζομένων είναι 1 έτος και η υψηλότερη 31 έτη. Το εύρος του χρόνου προϋπηρεσίας είναι 30 έτη και η τυπική απόκλιση 8,601. Στον παραπάνω πίνακα δίνονταιαναλυτικά όλες οι αριθμητικές τιμές, όπως προέκυψαν από τη στατιστική επεξεργασία των δεδομένων. Ακολουθεί και το θηκόγραμμα.



Έβδομη μεταβλητή: ΠΟΣΟΣΤΙΑΙΑ ΕΠΙΔΟΣΗ ΕΡΓΑΖΟΜΕΝΩΝ

Ποσοστιαία επίδοση εργαζομένου											
	N	Εύρος	Μικρότε-ρη τιμή	Μεγαλύ-τερη τιμή	Μέση τιμή	Τυπική απόκλιση	Διακύμ ανση	Λοξότητα		Κυρτότητα	
Ποσοσ. επίδοση	100	33	65	98	81,45	10,503	110,311	-,036	,241	-1,362	,478
N	**100**										

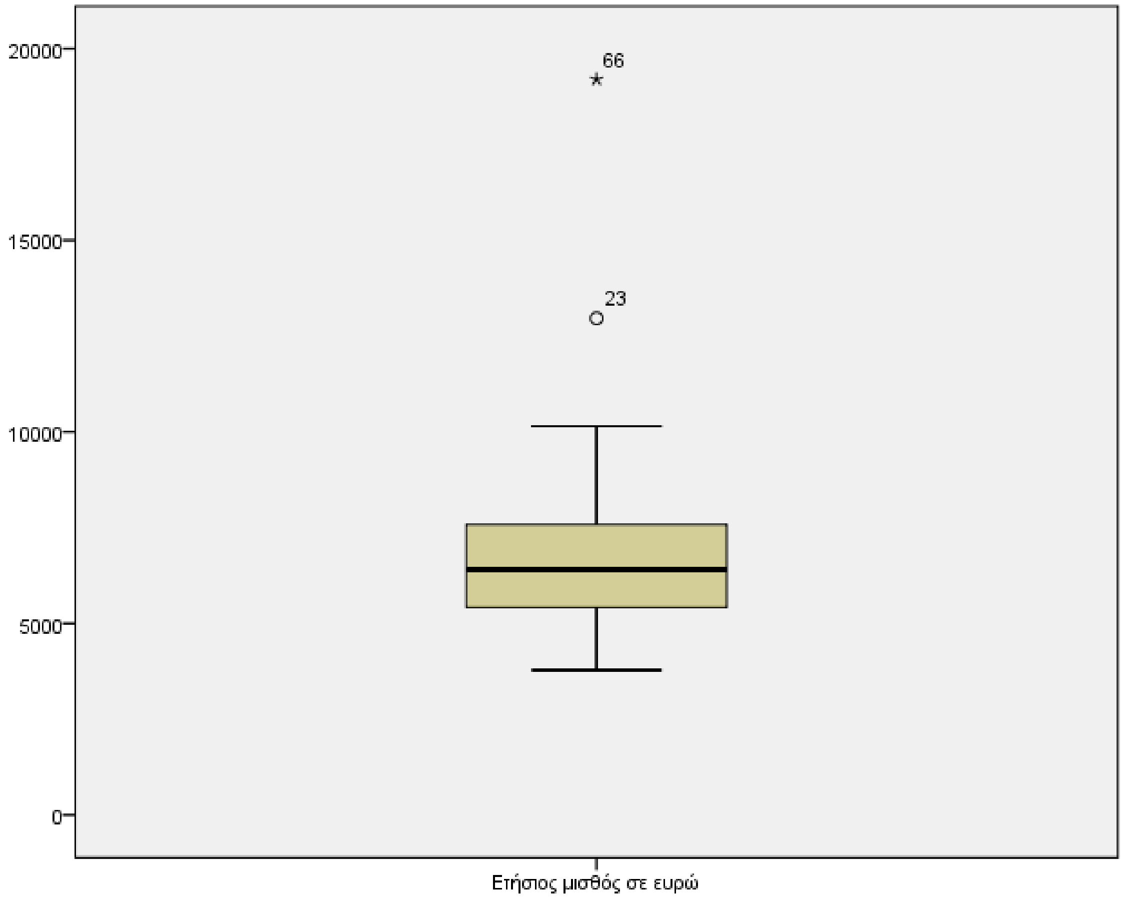
Η μέση ποσοστιαία επίδοση των εργαζομένων είναι 81,45%. Το χαμηλότερο ποσοστό επίδοσης είναι 65% και το υψηλότερο 98%. Το εύρος της συγκεκριμένης μεταβλητής είναι 33 και η τυπική απόκλιση 10,503.



Όγδοη μεταβλητή: ΕΤΗΣΙΟΣ ΜΙΣΘΟΣ

Ετήσιος μισθός σε ευρώ											
	Ν	Εύρος	Μικρότε-ρη τιμή	Μεγαλύ-τερη τιμή	Μέση τιμή	Τυπική απόκλιση	Διακύμανση	Λοξότητα		Κυρτότητα	
Ετήσιος μισθός	100	15420	3780	19200	6667,74	2007,678	4030771,204	2,796	,241	14,943	,47
N	**100**										

Ο μέσος ετήσιος μισθός των εργαζομένων της εταιρείας είναι 6.667,74 ευρώ. Ο μεγαλύτερος ετήσιος μισθός που δίνεται είναι 19.200 ευρώ και ο μικρότερος 3.780 ευρώ. Το εύρος του ετήσιου μισθού των εργαζομένων είναι 15.420 ευρώ και η τυπική απόκλιση 2.007,678. Ακολουθεί και το θηκόγραμμα, όπως προκύπτει από την ανάλυση της συγκεκριμένης μεταβλητής.



II. ΕΛΕΓΧΟΣ ΚΑΝΟΝΙΚΟΤΗΤΑΣ ΤΩΝ ΑΝΩΤΕΡΩ ΜΕΤΑΒΛΗΤΩΝ

Ο έλεγχος κανονικότητας των μεταβλητών της εργασίας έγινε με τα τεστ (Binomial, Chi - Square, Kolmogorov - Smirnov). Τα αποτελέσματα δίνονται αναλυτικά στον ακόλουθο πίνακα. Όπως διαφαίνεται, οι μεταβλητές Ηλικία, Ποσοστιαία επίδοση εργαζομένου και Ετήσιος μισός αποδέχονται την μηδενική υπόθεση και οι υπόλοιπες απορρίπτουν την μηδενική υπόθεση.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The categories defined by Φύλο = άνδρας and γυναίκα occur with probabilities 0,5 and 0,5.	One-Sample Binomial Test	,012	Reject the null hypothesis.
2	The categories of Κατηγορία εργασίας occur with equal probabilities.	One-Sample Chi-Square Test	,000	Reject the null hypothesis.
3	The categories of Σκοπός σπουδών occur with equal probabilities.	One-Sample Chi-Square Test	,004	Reject the null hypothesis.
4	The distribution of Ηλικία is normal with mean 37,840 and standard deviation 10,09.	One-Sample Kolmogorov-Smirnov Test	,091	Retain the null hypothesis.
5	The distribution of Χρόνια εκπαίδευσης is normal with mean 12,480 and standard deviation 2,43	One-Sample Kolmogorov-Smirnov Test	,000	Reject the null hypothesis.
6	The distribution of Χρόνος προϋπηρεσία is normal with mean 9,480 and standard deviation 8,60.	One-Sample Kolmogorov-Smirnov Test	,000	Reject the null hypothesis.
7	The distribution of Ποσοστιαία επίδοση εργαζομένου is normal with mean 81,450 and standard deviation 10,50.	One-Sample Kolmogorov-Smirnov Test	,247	Retain the null hypothesis.
8	The distribution of Ετήσιος μισθός σε ευρώ is normal with mean 6.667,740 and standard deviation 2.007,68.	One-Sample Kolmogorov-Smirnov Test	,163	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is ,05.

III. ΕΛΕΓΧΟΣ ΥΠΟΘΕΣΕΩΝ

Πραγματοποιήθηκαν **One sample T – Test** (τεστ έλεγχου υποθέσεων μιας μεταβλητής) στις μεταβλητές που αναλύονται παρακάτω.

One-Sample T: Ηλικία

Σε αυτό το σημείο πραγματοποιείται έλεγχος υπόθεσης σχετικά με την ηλικία. Έστω, ότι η μηδενική υπόθεση Η0: μ = 35 ετών και η εναλλακτική υπόθεση Η1: μ > 35. Ελέγχεται εάν η μέση ηλικία είναι 35 ετών ή αν είναι μεγαλύτερη από 35 ετών και ανάλογα απορρίπτουμε ή αποδεχόμαστε τη μηδενική υπόθεση. Σύμφωνα με τη θεωρία, εάν η μέση ηλικία είναι 35 την αποδεχόμαστε, διαφορετικά την απορρίπτουμε. Ακολουθούν τα αποτελέσματα όπως προκύπτουν από τη στατιστική επεξεργασία των δεδομένων.

Test of mu = 35 vs > 35

Variable	N	Mean	St.Dev.	SE Mean	*95%* Lower Bound	T	P
Ηλικία	100	37,84	10,0862	1,0086	36,1653	2,82	0,003

Όπως προκύπτει, η Η0 απορρίπτεται γιατί η μέση ηλικία είναι μεγαλύτερη από 35 (μ = 37,84 ετών). Το επίπεδο σημαντικότητας είναι 5%, επειδή P (p – value) = 0,003 < 0,05.

One-Sample T: Χρόνια εκπαίδευσης

Εν συνεχεία πραγματοποιείται έλεγχος υπόθεσης σχετικά με τα χρόνια εκπαίδευσης. Έστω, ότι η μηδενική υπόθεση H0: μ = 10 έτη και η εναλλακτική υπόθεση H1: μ > 10 ετών. Στον πίνακα που ακολουθεί δίνονται τα αποτελέσματα του έλεγχου υποθέσεων της μεταβλητής αυτής.

Test of mu = 10 vs > 10

			95%				
			Lower				
Variable	N	Mean	St.Dev.	SE Mean	Bound	T	P
Χρόνια εκπαιδ.	100	12,48	2,4349	0,2435	12,0757	10,19	0,000

Όπως προκύπτει, η H0 απορρίπτεται γιατί η μέση τιμή των ετών εκπαίδευσης είναι μεγαλύτερη από 10 (μ = 12,48 έτη). Το επίπεδο σημαντικότητας είναι 5%, επειδή P (p – value) = 0,000 < 0,05.

One-Sample T: Επίδοση

Στον ακόλουθο πίνακα δίνονται τα αποτελέσματα του έλεγχου υποθέσεων της μεταβλητής επίδοση ποσοστιαία των εργαζομένων (δείγμα). Έστω, ότι η μηδενική υπόθεση H0: μ = 50 και η εναλλακτική υπόθεση H1: μ > 50. Η μηδενική υπόθεση απορρίπτεται και σε αυτήν την μεταβλητή.

Test of mu = 50 vs > 50

Variable	N	Mean	St.Dev.	SE Mean	95% Lower Bound	T	P
Επίδοση	100	81,45	10,5029	1,0503	79,7061	29,94	0,000

Το επίπεδο σημαντικότητας είναι 5%, επειδή P (p – value) = 0,000 < 0,05.

One-Sample T: Μισθός

Στον πίνακα που ακολουθεί παρουσιάζονται τα αποτελέσματα του έλεγχου υποθέσεων του μισθού των εργαζομένων. Έστω, ότι η μηδενική υπόθεση H0: μ = 7000 ευρώ και η εναλλακτική υπόθεση H1: μ > 7000 ευρώ. Η μηδενική υπόθεση δεν απορρίπτεται αφού ο μέσος μισθός του δείγματος είναι 6667,74. Η εναλλακτική υπόθεση δηλώνει ότι ο μέσος μισθός είναι μεγαλύτερος από 7000 ευρώ και στην συγκεκριμένη περίπτωση είναι μικρότερος (H0 ≠ H1). Το P (p – value) = 0,949 > 0,05, δηλαδή το επίπεδο σημαντικότητας είναι χαμηλό.

Test of mu = 7000 vs > 7000

Variable	N	Mean	St.Dev.	SE Mean	95% Lower Bound	T	P
Μισθός (ευρώ)	100	6667,74	2007,68	200,77	6334,39	-1,65	0,949

One-Sample T: Χρόνια εκπαίδευσης

Σε αυτό το σημείο γίνεται έλεγχος υπόθεσης σχετικά με τα χρόνια εκπαίδευσης, αλλά με άλλη μέση τιμή ετών εκπαίδευσης από πριν. Έστω, ότι η μηδενική υπόθεση H0: μ = 12 έτη και η εναλλακτική υπόθεση H1: μ < 12 ετών. Στον πίνακα που ακολουθεί δίνονται τα αποτελέσματα του έλεγχου υποθέσεων της μεταβλητής αυτής. Η μηδενική υπόθεση δεν απορρίπτεται. Η εναλλακτική υπόθεση δηλώνει ότι η μέση τιμή των ετών εκπαίδευσης πρέπει να είναι μικρότερη από τα 12 έτη και στην συγκεκριμένη περίπτωση είναι μεγαλύτερη (H0 ≠ H1).

Test of mu = 12 vs < 12

				95%			
				Upper			
Variable	N	Mean	St.Dev.	SE Mean	Bound	T	P
Χρόνια εκπαιδ.	100	12,48	2,4349	0,2435	12,8843	1,97	0,974

One-Sample T: Προϋπηρεσία

Στον ακόλουθο πίνακα δίνονται τα αποτελέσματα του έλεγχου υποθέσεων την προϋπηρεσίας των εργαζομένων. Έστω, ότι η μηδενική υπόθεση H0: μ = 10 έτη προϋπηρεσίας και η εναλλακτική υπόθεση H1: μ > 10 έτη. Η μηδενική υπόθεση δεν απορρίπτεται.

Test of mu = 10 vs > 10

					95%		
					Lower		
Variable	N	Mean	St.Dev.	SE Mean	Bound	T	P
Προϋπηρεσία	100	9,48	8,60054	0,86005	8,05197	-0,60	0,727

Επίσης πραγματοποιούνται **Two – Sample T-Test** (έλεγχος υποθέσεων για δύο δείγματα). Στη συνέχεια παρουσιάζονται και αναλύονται τα αποτελέσματα των ελέγχων αυτών.

Two-Sample T-Test and CI: Ηλικία; Φύλο

Στην συγκεκριμένη περίπτωση έγινε έλεγχος υποθέσεων των μεταβλητών ηλικία και φύλο. Αρχικά γίνεται έλεγχος μέσης ηλικίας δείγματος και συγκεκριμένα ελέγχεται εάν η μέση ηλικία ανδρών είναι ίση με τη μέση ηλικία γυναικών. Ισχύει η μηδενική υπόθεση Η0: μ1 = μ2 και η εναλλακτική υπόθεση Η1: μ1 ≠ μ2. Ακολουθούν πίνακες με τα αποτελέσματα.

Two-sample T for Ηλικία

Φύλο	N	Mean	StDev	SE
1	**37**	**35,27**	**9,06**	**1,5**
2	**63**	**39,3**	**10,4**	**1,3**

Difference = mu (1) - mu (2)
Estimate for difference: -4,07894
95% CI for difference: (-8,16475; 0,00688)
T-Test of difference = 0 (vs not =): T-Value = -1,98 P-Value = 0,050 **DF = 98**
Both use Pooled St.Dev. = 9,9404

Από τον συγκεκριμένο έλεγχο προκύπτει ότι δεν υπάρχει διαφορά στην ηλικία ανά φύλο. Συνεπώς, η μηδενική υπόθεση δεν απορρίπτεται.

Test for Equal Variances: Ηλικία versus Φύλο

Σε αυτό το σημείο γίνεται σύγκριση των διασπορών (δηλαδή ελέγχεται εάν οι διασπορές είναι ίσες) δυο μεταβλητών, της ηλικίας και του φύλου. Τα αποτελέσματα δίνονται στους πίνακες που ακολουθούν. Επίσης, γίνεται σύγκριση διασπορών των μεταβλητών ηλικία και χρόνια εκπαίδευσης. Από τους ελέγχους (**F Test και Levene's Test**) προκύπτει ότι η διασπορά των μεταβλητών που συγκρίθηκα είναι ίσες.

95% Bonferroni confidence intervals for standard deviations

Φύλο	N	Lower	StDev	Upper
1	37	7,16	9,06	12,25
2	63	8,66	10,4	13,00

F-Test (normal distribution)
Test statistic = 0,76; p-value = 0,371
Levene's Test (any continuous distribution)
Test statistic = 1,61; p-value = 0,207

Test for Equal Variances for Ηλικία

Test for Equal Variances: Χρόνια εκπαίδευσης versus Φύλο

95% Bonferroni confidence intervals for standard deviations

Φύλο	N	Lower	StDev	Upper
1	37	1,82	2,30	3,11
2	63	2,00	2,41	3,01

F-Test (normal distribution)
Test statistic = 0,91; p-value = 0,767
Levene's Test (any continuous distribution)

Test statistic = 0,23; p-value = 0,632

Test for Equal Variances for Χρόνια εκπαίδευσης

Two-Sample T-Test and CI ; Φύλο

Στη συνέχεια γίνεται έλεγχος δυο δειγμάτων T Test των μεταβλητών φύλο και χρόνια εκπαίδευσης. Προκύπτει ότι το φύλο διαφέρει από τα χρόνια εκπαίδευσης (H1≠H0), συνεπώς απορρίπτεται η μηδενική υπόθεση. Το p – value (P) ισούται με 0,016 < 0,05. Ακολουθούν δύο πίνακες που παρουσιάζουν αναλυτικά τα αποτελέσματα των ελέγχων αυτών.

Two-sample T for Χρόνια εκπαίδευσης

Φύλο	N	Mean	StDev	SE
1	37	13,24	2,30	0,38
2	63	12,03	2,42	0,30

Difference = mu (1) - mu (2)
Estimate for difference: 1,21150
95% CI for difference: (0,23533; 2,18766)
T-Test of difference = 0 (vs not =): T-Value = 2,46 P-Value = 0,016 DF = 98
Both use Pooled StDev = 2,3749

Two-Sample T-Test and CI: Προϋπηρεσία; Φύλο

Επίσης, γίνεται έλεγχος δυο δειγμάτων T Test των μεταβλητών φύλο και έτη προϋπηρεσίας. Σύμφωνα με τα αποτελέσματα η μηδενική υπόθεση δεν απορρίπτεται, αφού η προϋπηρεσία δεν διαφοροποιείται ως προς το φύλο. Το p – value (P) ισούται με 0,366 > 0,05. Ακολουθούν δύο πίνακες που παρουσιάζουν αναλυτικά τα αποτελέσματα των ελέγχων αυτών.

Two-sample T for Προϋπηρεσία

Φύλο	**N**	**Mean**	**StDev**	**SE**
1	**37**	**8,46**	**8,07**	**1,3**
2	**63**	**10,08**	**8,91**	**1,1**

Difference = mu (1) - mu (2)
Estimate for difference: -1,61991
95% CI for difference: (-5,15810; 1,91829)
T-Test of difference = 0 (vs not =): T-Value = -0,91 P-Value = 0,366 **DF = 98**
Both use Pooled StDev = 8,6081

Two-Sample T-Test and CI: Επίδοση; Φύλο

Πραγματοποιείται και έλεγχος δυο δειγμάτων T Test των μεταβλητών φύλο και επίδοση. Σύμφωνα με τα αποτελέσματα η μηδενική υπόθεση δεν απορρίπτεται, αφού η επίδοση των εργαζομένων δεν διαφοροποιείται ως προς το φύλο τους. Το p – value (P) ισούται με 0,213 > 0,05. Ακολουθούν δύο πίνακες που παρουσιάζουν αναλυτικά τα αποτελέσματα των ελέγχων αυτών.

Two-sample T for Επίδοση

Φύλο	**N**	**Mean**	**StDev**	**SE**
1	**37**	**83,2**	**11,2**	**1,8**
2	**63**	**80,4**	**10,0**	**1,3**

Difference = mu (1) - mu (2)
Estimate for difference: 2,71772
95% CI for difference: (-1,58691; 7,02234)
T-Test of difference = 0 (vs not =): T-Value = 1,25 P-Value = 0,213 **DF = 98**
Both use Pooled StDev = 10,4728

Two-Sample T-Test and CI: Μισθός; Φύλο

Τέλος, γίνεται έλεγχος δυο δειγμάτων T Test των μεταβλητών φύλο και μισθός. Προκύπτει ότι ο μισθός διαφέρει ως προς το φύλο των εργαζομένων που συμμετείχαν στην έρευνα, συνεπώς απορρίπτεται η μηδενική υπόθεση. Το p – value (P) ισούται με 0,000 < 0,05. Ακολουθούν πίνακες με αναλυτικά αποτελέσματα των ελέγχων αυτών.

Two-sample T for Μισθός

Φύλο	N	Mean	StDev	SE Mean
1	**37**	**7604**	**2279**	**375**
2	**63**	**6118**	**1610**	**203**

Difference = mu (1) - mu (2)
Estimate for difference: 1485,96
95% CI for difference: (711,88; 2260,03)
T-Test of difference = 0 (vs not =): T-Value = 3,81 P-Value = 0,000 **DF = 98**
Both use Pooled StDev = 1883,2679

ΣΥΜΠΕΡΑΣΜΑΤΑ

Τα διαστήματα εμπιστοσύνης, σε συνδυασμό με τους ελέγχους των υποθέσεων, αποτελούν το σημαντικότερο κομμάτι της επιστήμης της στατιστικής συμπερασματολογίας.

Ο έλεγχος των υποθέσεων στην οικονομική επιστήμη εφαρμόζεται κυρίως μέσα από την οικονομετρική ανάλυση. Η οικονομετρία αποτελεί ένα σύγχρονο εργαλείο ανάλυσης των οικονομικών, αλλά και των χρηματοοικονομικών μεγεθών.

Ο έλεγχος μιας υπόθεσης παρέχει πληροφορίες στον ενδιαφερόμενο, για τα δείγματα του πληθυσμού για τα οποία η υπόθεση H_0 γίνεται αποδεκτή, καθώς επίσης και για τα δείγματα για τα οποία η υπόθεση απορρίπτεται.

Ουσιαστικά, ένας έλεγχος υπόθεσης διαιρεί το χώρο σε δύο αμοιβαία αποκλειόμενα υποσύνολα, στο ένα από τα οποία αποδεχόμαστε την υπόθεση H_0 και στο άλλο στο οποίο η μηδενική υπόθεση απορρίπτεται. Συμπερασματικά, αναφέρεται πως στην παρούσα μελέτη περίπτωσης έγινε εφαρμογή της θεωρίας του έλεγχου υποθέσεων, που αναφέρθηκε σε προηγούμενα κεφάλαια. Αρχικά, πραγματοποιήθηκε περιγραφική ανάλυση των μεταβλητών της έρευνας, δηλαδή μονομεταβλητή ανάλυση και επίσης, παρατέθηκαν και τα θηκογράμματα (boxplot) των μεταβλητών αυτών. Οι μεταβλητές που αναλύθηκαν ήταν οχτώ. Στη συνέχεια έγινε έλεγχος κανονικότητας των οχτώ αυτών μεταβλητών.

Τέλος, πραγματοποιήθηκε έλεγχος υποθέσεων ενός δείγματος, δύο δειγμάτων και σύγκριση διασπορών. Ο έλεγχος υποθέσεων ενός δείγματος (One Sample T – Test) έγινε στις μεταβλητές ηλικία, χρόνια εκπαίδευσης, επίδοση, μισθός και προϋπηρεσία και σε κάποιες περιπτώσεις απορρίφτηκε η μηδενική υπόθεση, σε κάποιες άλλες όχι. Ο έλεγχος υποθέσεων δύο δειγμάτων (Two Sample T – Test) έγινε στις μεταβλητές ηλικία, επίδοση, χρόνια εκπαίδευσης και μισθός, ως προς το φύλο.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Χατζηνικολάου Δ. (2002). «Στατιστική για οικονομολόγους. Β' Έκδοση». Εκδόσεις Printshop, Θεσσαλονίκη

Ψώνιος Δ., (1999). «Στατιστική». Εκδόσεις Ζήτη, Θεσσαλονίκη

Ρούσσας Γ. (1994). «Στατιστική συμπερασματολογία Τ. ΙΙ, Έλεγχος υποθέσεων». Εκδόσεις Ζήτη, Αθήνα

Begg David, Fischer Stanley, Dornbusch Rudiger (2006). «Εισαγωγή στην οικονομική». Εκδόσεις Κριτική, Αθήνα

Τζαβαλής Ηλίας, (2008). «Οικονομετρία». Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών

Βενέτης Ιωάννης, (2009). «Εισαγωγικές διαλέξεις στην Οικονομετρία». Εκδόσεις Γκιούρδας, Αθήνα

Κιντής Α., (2010). «Σύγχρονη οικονομετρική ανάλυση - Τόμος Α'». Εκδόσεις Gutenberg, Αθήνα

Αγιακλόγλου Χρήστος, Μπένος Θεοφάνης, (2007). «Εισαγωγή στην οικονομετρική ανάλυση». Εκδόσεις Μπένου

Τσιώνας Βασίλειος, (2009). «Στατιστική με εφαρμογές στην Οικονομετρία». Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών

Βάμβουκας Γεώργιος, (2007). «Σύγχρονη Οικονομετρία». Εκδόσεις Οικονομικού Πανεπιστημίου Αθηνών

Κολύβα-Μαχαίρα Φ., Μπόρα-Σέντα Ε., (1998). «Στατιστική θεωρία και εφαρμογές». Εκδόσεις Ζήτη, Θεσσαλονίκη

Τριχόπουλος Δ., Τζώνου Α., Κατσουγιάννη κ., (2001). «Βιοστατιστική». Εκδόσεις Παρισιάνου, Αθήνα

Ατζουλάκος Δ., (2008). «Στατιστικός Έλεγχος Ποιότητας – Β’ Έκδοση». Σημειώσεις παραδόσεων. Τμήμα Στατιστικής και Ασφαλιστικής Επιστήμης, Πανεπιστήμιο Πειραιώς

ΔΙΑΔΙΚΤΥΑΚΕΣ ΠΗΓΕΣ

http://www.stat-athens.aueb.gr/~jpan/statistiki-skepsi-II/chapter21.pdf

http://www.aua.gr/gpapadopoulos/files/hypoth-tests-4.pdf

http://users.auth.gr/dkugiu/Teach/CivilEngineer/hypothesis.pdf

http://www.mednet.gr/archives/2010-4/pdf/691.pdf

http://www.aua.gr/gpapadopoulos/files/anova12-13a.pdf

ΠΑΡΑΡΤΗΜΑ

Σας παραθέτω τα δεδομένα που θα χρησιμοποιηθούν σε αυτή τη μελέτη περίπτωσης αναλυτικά παρακάτω:

ΜΕΤΑΒΛΗΤΕΣ:

• **ΦΥΛΟ (SEX):** 1 = ΑΝΔΡΑΣ 2 = ΓΥΝΑΙΚΑ
• **ΗΛΙΚΙΑ (AGE):** …….ΣΕ ΕΤΗ
• **ΚΑΤΗΓΟΡΙΑ ΕΡΓΑΣΙΑΣ (CATEG_ERG)**: 1 = ΥΠΑΛΛΗΛΟΣ ΓΡΑΦΕΙΟΥ, 2 = ΕΚΠΑΙΔΕΥΟΜΕΝΟΣ ΥΠΑΛΛΗΛΟΣ, 3 = ΑΣΦΑΛΕΙΑ, 4 = ΠΤΥΧΙΟΥΧΟΣ, 5 = ΠΤΥΧΙΟΥΧΟΣ ΜΕ Μ.Δ.Ε, 6 = ΧΡΗΜΑΤ. ΑΝΑΛΥΤΗΣ, 7 = ΠΡΟΙΣΤΑΜΕΝΟΣ
• **ΧΡΟΝΙΑ ΕΚΠΑΙΔΕΥΣΗΣ (EDUC_YEARS**): ………..ΣΕ ΕΤΗ
• **ΣΚΟΠΟΣ ΣΠΟΥΔΩΝ (SKOPOS_SPOUDON**):1 = ΛΗΨΗ ΔΙΠΛΩΜΑΤΟΣ, 2 = ΑΠΟΚΤΗΣΗ ΓΝΩΣΕΩΝ, 3 = ΑΝΑΒΟΛΗ ΣΤΡΑΤΕΥΣΗΣ, 4= ΕΥΡΕΣΗ ΕΡΓΑΣΙΑΣ
• **ΠΡΟΫΠΗΡΕΣΙΑ (PROIPIRESIA)** : ΧΡΟΝΟΣ ΠΡΟΥΠΗΡΕΣΙΑΣ ΣΕ ΕΤΗ
• **ΕΠΙΔΟΣΗ (EPIDOSI)** : ΠΟΣΟΣΤΙΑΙΑ ΕΠΙΔΟΣΗ ΕΡΓΑΖΟΜΕΝΟΥ
• **ΜΙΣΘΟΣ (ETISIOS_MISTHOS) :** ΕΤΗΣΙΟΣ ΜΙΣΘΟΣ ΣΕ ΕΥΡΩ

Σας παραθέτω και τον πίνακα με τα δεδομένα κωδικοποιημένα πριν εισαχθούν στο στατιστικό πακέτο για επεξεργασία:

ΠΙΝΑΚΑΣ ΔΕΔΟΜΕΝΩΝ:

sex	age	categ_erg	educ_years	skopos_spoudon	proipiresia	epidosi	etisios_misthos
1	30	1	15	4	6	66	9.684
1	39	1	12	3	11	87	6.192
1	30	1	15	2	5	77	8.280
1	30	1	8	2	4	95	7.020
1	60	1	8	3	22	66	6.300
1	46	1	14	1	20	78	6.228

1	25	1	12	4	4	88	7.416
1	30	1	12	4	2	78	6.408
1	32	6	12	1	6	97	10.152
1	35	1	12	2	8	89	6.552
1	38	1	15	3	6	94	6.444
1	42	1	8	3	14	85	6.300
1	36	1	15	2	2	91	7.596
1	30	1	15	1	4	98	7.704
1	29	1	15	1	3	66	7.776
1	25	1	15	4	4	98	8.640
1	33	1	15	4	5	89	6.480
1	30	1	15	4	4	86	7.416
1	28	1	8	3	3	80	5.112
2	29	1	12	2	6	98	5.148
2	60	1	8	1	30	88	4.824
2	55	1	12	4	25	76	5.292
2	33	7	16	4	1	93	12.960
2	44	1	15	1	2	86	7.992
2	39	1	12	1	8	82	5.760
2	38	1	12	2	5	81	6.660
2	60	1	12	2	20	85	6.840
2	32	1	12	4	7	84	8.640
2	62	1	8	4	22	70	3.780
2	48	1	8	1	16	92	3.888
2	43	1	12	1	15	75	5.868
2	38	1	8	2	6	70	4.716
2	36	1	12	2	8	90	7.596
2	35	1	8	4	5	92	5.760
2	28	1	12	4	3	69	5.508
2	26	1	8	4	1	72	4.248
2	38	1	8	2	3	74	3.816
2	41	1	8	2	10	65	5.184
2	45	1	12	1	15	66	7.128
2	54	1	12	1	23	70	5.040
2	58	1	12	1	26	67	5.328
2	35	1	12	1	10	73	5.544
2	30	1	12	2	4	80	4.788
2	27	1	15	1	1	82	5.256
2	34	1	8	2	2	74	5.688
2	32	1	12	4	3	97	6.300
2	26	1	12	1	1	66	5.868
2	45	1	12	4	20	87	6.408
2	53	1	12	2	30	91	5.040
2	44	1	12	4	14	88	6.840
2	46	5	16	4	20	93	8.352

2	30	1	12	1	1	98	5.760
2	31	1	12	1	3	86	4.356
2	29	1	16	2	3	97	9.324
2	27	1	15	2	1	90	4.500
2	26	1	12	4	2	73	9.036
2	30	1	8	2	4	83	4.752
2	43	1	12	1	6	84	3.960
2	58	1	15	1	12	82	6.156
2	50	1	12	2	18	89	6.300
2	40	1	8	4	14	72	4.068
1	30	2	12	3	2	70	5.400
1	39	2	12	1	10	70	6.588
1	30	2	15	2	3	96	7.812
1	30	2	15	4	1	72	6.804
1	60	7	15	3	28	82	19.200
1	46	4	15	4	24	67	7.488
1	25	2	15	1	1	79	6.660
1	30	2	15	2	7	92	7.596
1	32	2	15	4	3	83	7.956
1	35	2	12	4	2	95	5.580
1	38	6	12	3	12	96	8.676
1	41	5	12	1	19	66	7.264
1	38	6	15	2	11	96	10.080
1	25	2	15	3	1	70	7.164
1	26	2	12	4	1	94	6.480
1	55	2	15	4	30	65	7.596
1	33	2	12	3	6	82	7.164
1	44	2	15	2	19	94	8.136
2	39	2	12	1	12	65	5.220
2	38	4	15	4	6	69	5.796
2	60	5	16	4	28	85	8.568
2	32	2	12	2	2	69	5.184
2	62	2	12	1	31	66	4.896
2	48	2	12	1	19	93	6.336
2	43	2	12	2	12	73	5.328
2	38	2	8	2	10	69	6.984
2	36	2	15	4	7	93	5.796
2	35	2	12	4	5	69	5.436
2	28	2	12	4	3	90	7.560
2	30	6	15	4	2	93	7.020
2	33	2	12	2	3	80	6.840
2	39	4	15	2	3	88	8.532
2	46	2	12	1	22	74	6.264
2	43	2	15	1	24	76	6.120
2	29	2	15	2	2	73	5.508

2	25	2	12	4	1	93	5.508
2	32	5	16	4	8	75	7.920
2	30	2	12	4	4	76	6.998
2	35	2	12	4	5	69	7.344